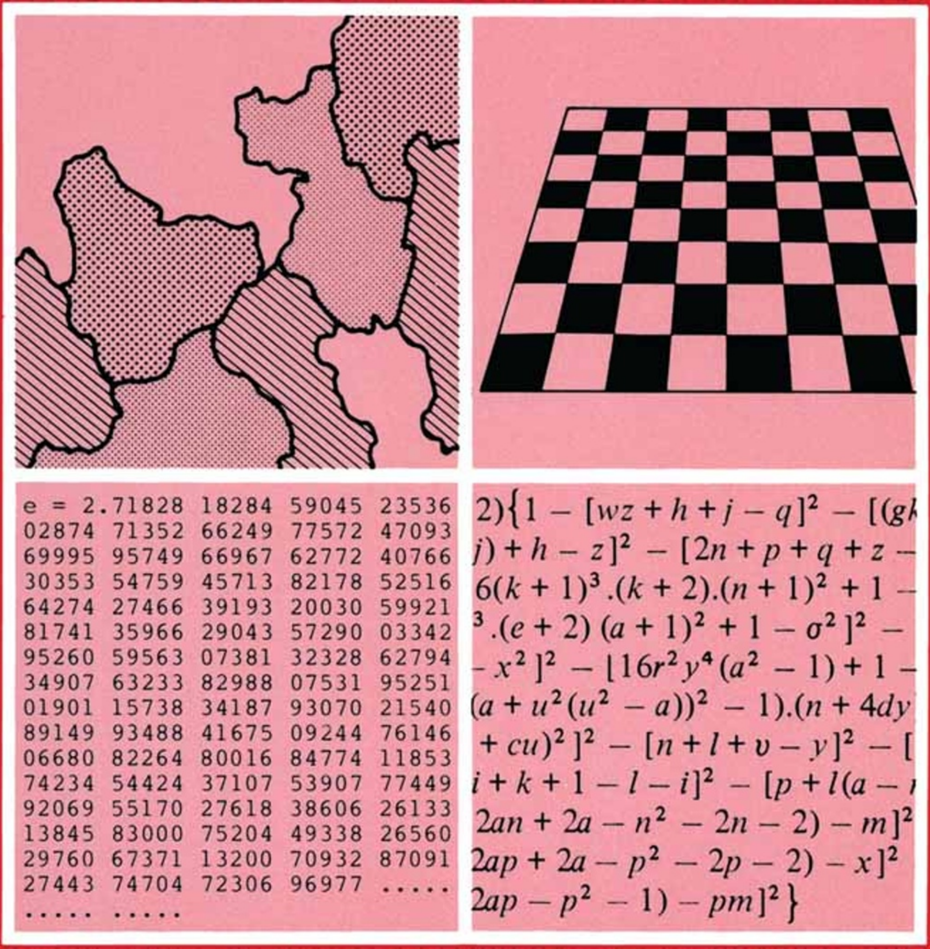
# Micro-Maths

## Mathematical problems and theorems to consider and solve on a computer


e = 2.71828 18284 59045 23536
02874 71352 66249 77572 47093
69995 95749 66967 62772 40766
30353 54759 45713 82178 52516
64274 27466 39193 20030 59921
81741 35966 29043 57290 03342
95260 59563 07381 32328 62794
34907 63233 82988 07531 95251
01901 15738 34187 93070 21540
89149 93488 41675 09244 76146
06680 82264 80016 84774 11853
74234 54424 37107 53907 77449
92069 55170 27618 38606 26133
13845 83000 75204 49338 26560
29760 67371 13200 70932 87091
27443 74704 72306 96977 .....
..... .....
2){1 − [wz + h + j − q]² − [(g
j) + h − z]² − [2n + p + q + z −
6(k + 1)³.(k + 2).(n + 1)² + 1 −
³.(e + 2) (a + 1)² + 1 − σ²]² −
− x²]² − [16r²y⁴(a² − 1) + 1 −
(a + u²(u² − a))² − 1).(n + 4dy
+ cu)²]² − [n + l + v − y]² − [
i + k + 1 − l − i]² − [p + l(a −
2an + 2a − n² − 2n − 2) − m]²
2ap + 2a − p² − 2p − 2) − x]²
2ap − p² − 1) − pm]²}

KEITH DEVLIN

# Micro-Maths

**Other Macmillan titles of related interest**

*Advanced Graphics with the Acorn Electron*
Ian O. Angell and Brian J. Jones
*Advanced Graphics with the BBC Microcomputer*
Ian O. Angell and Brian J. Jones
*Advanced Graphics with the Sinclair ZX Spectrum*
Ian O. Angell and Brian J. Jones
*Assembly Language Programming for the Acorn Electron*
Ian Birnbaum
*Assembly Language Programming for the BBC Microcomputer*, second edition
Ian Birnbaum
*Advanced Programming for the 16K ZX81* Mike Costello
*Using Your Home Computer* Garth W. P. Davies
*Beginning BASIC* Peter Gosling
*Continuing BASIC* Peter Gosling
*Practical BASIC Programming* Peter Gosling
*Program Your Microcomputer in BASIC* Peter Gosling
*Codes for Computers and Microprocessors* P. Gosling and Q. Laarhoven
*Microprocessors and Microcomputers – their use and programming*
Eric Huggins
*The Sinclair ZX81 – Programming for Real Applications*
Randle Hurley
*More Real Applications for the ZX81 and ZX Spectrum*
Randle Hurley
*Programming in Z80 Assembly Language* Roger Hutty
*Beginning BASIC with the ZX Spectrum* Judith Miller
*Digital Techniques* Noel Morris
*Microprocessor and Microcomputer Technology* Noel Morris
*Using Sound and Speech on the BBC Microcomputer*
M. A. Phillips
*The Alien, Numbereater, and Other Programs for Personal Computers – with notes on how they were written* John Race
*Understanding Microprocessors* B. S. Walker
*Assembly Language Assembled – for the Sinclair ZX81*
Anthony Woods

# Micro-Maths

## Mathematical problems and theorems to consider and solve on a computer

Keith Devlin

MACMILLAN

© Keith Devlin 1984

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright Act 1956 (as amended).

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

First published 1984

Published by
MACMILLAN PUBLISHERS LTD
Houndmills, Basingstoke, Hampshire RG21 2XS
and London
Companies and representatives
throughout the world

British Library Cataloguing in Publication Data
Devlin, Keith
Micro-maths: mathematical problems and theorems
to consider and solve on a computer.
1. Mathematics–Data processing 2. Microcomputers
I. Title
510'.28'5404 QA76.95

ISBN 978-1-349-07938-4 ISBN 978-1-349-07936-0 (eBook)
DOI 10.1007/978-1-349-07936-0

# *Contents*

*About this book* *vii*

*Acknowledgements* *ix*

*About the author* *xi*

*The first problem* *xiii*

1 Computer mathematics reaches its prime 1
2 Pi and chips 11
3 Formulas for primes 17
4 The kilderkin approach through a silicon gate 23
5 Colouring by numbers 29
6 The Oxen of the Sun (or how Archimedes' number came up 2000 years too late) 35
7 100 year old problem solved 41
8 Mod mathematics 1801 style 47
9 Another slice of pi 53
10 Coincidence? 59
11 Fermat's Last Theorem 65
12 Seven-up 71
13 Primes and secret codes 79
14 Perfect numbers 87
15 True beyond reasonable doubt 93
16 All numbers great and small 99

*Table of the Mersenne primes known in June 1984* *102*

*Crib* *103*

# *About this book*

All of the articles and problems in this book first appeared in *The Guardian* newspaper during the years 1983 and 1984, though in many cases I have extended the necessarily brief accounts originally given, and on some occasions I have amalgamated two articles into one chapter.

As with my *Guardian* column, there is no particular connection between one chapter and the next. By and large, you should be able to pick up the book and delve into it at random. There is no overall theme, save that everything concerns computing and mathematics. The choice of the items chosen was a simple one: I write about whatever I find fun and of interest. If your favourite topic is not here, drop me a line and tell me about it, and I will see if I can include it in a future column (or even a future edition of this book).

*Lancaster University* Keith Devlin
*August 1984*

# *Acknowledgements*

The book is dedicated to my two editors at *The Guardian*, Tim Radford and Anthony Tucker, for giving me the opportunity to spout off to an audience somewhat larger than the one usually provided for me. Is there another national daily newspaper in the world which would devote a regular column to mathematics?

# *About the author*

Dr Keith Devlin is Reader in Mathematics at The University of Lancaster. Since the spring of 1983 he has written occasional articles on mathematics and computing in *The Guardian* newspaper, and has contributed a regular, fortnightly column to the computer page ('*Microguardian*') since it began in the autumn of 1983.

In addition to this book, he has written half a dozen other mathematics books, most of them dry old textbooks destined to accumulate dust in obscure corners of university libraries.

Confirming the popular impression that you have to be a masochist to enjoy mathematics, his main interest outside of the subject is fell-running, an interest not shared by his wife and two children, who are content to merely *look* at the fells from their house in the Lune Valley in Lancashire.

# *The first problem*

## *100 Sums*

If you take the digits 1 to 9 in order, there are exactly 11 ways in which you can insert plus and minus signs to give a sum with answer 100. One of these is

$$123 - 45 - 67 + 89 = 100$$

Find the other 10.

This problem is a good one for computer attack, though the patient among you could presumably get it out using nothing more high-tech than paper and a pencil.

# *1 Computer mathematics reaches its prime*

A positive whole number $N$ is called a *prime* number if the only whole numbers which divide exactly into it are 1 and $N$ itself. For example, of the first twenty numbers, 2, 3, 5, 7, 11, 13, 17, 19 are primes whereas 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20 are not. (The number 1 is conventionally excluded from the category of primes.) Except for the number 2, all primes are odd – a fact which makes 2 a very 'odd' prime, of course. But there are plenty of odd numbers which are not prime; for instance 9, 15, 81.

An important property of (positive, whole) numbers is expressed by what mathematicians call 'The Fundamental Theorem of Arithmetic'. This says that every number other than 1 is either prime itself or else can be written as a product of two or more prime numbers. Furthermore, any expression of a number as a product of prime numbers is unique except for a possible rearranging of the primes involved. For example, $6 = 2 \times 3$, $21 = 3 \times 7$, $84 = 2 \times 2 \times 3 \times 7$. This fact means that the primes can be regarded as the 'building blocks' out of which all whole numbers are constructed. Indeed, the area of mathematics known as 'Number Theory', which deals with the properties of the whole numbers, is very largely concerned with the properties of the primes. It is not much of an exaggeration to say that if you understand all there is to know about the primes, then you understand everything about all whole numbers. Not that mathematicians do know all there is to know about the primes: in this book you will come across several examples of simple questions about primes which have not yet been resolved, even after centuries of effort.

The Fundamental Theorem of Arithmetic mentioned above was probably known to the Ancient Greek mathematicians who followed the teachings of Pythagoras, around 500 B.C. They seem to be the first to have studied the concept of prime numbers. Certainly the result appeared in Euclid's famous mathematics textbook *Elements*, written around 350 B.C. Also in Euclid's *Elements*, it was shown that there are an infinite number of primes. The demonstration of this fact remains to this day a wonderful example (albeit a very simple one) of what it takes to constitute a 'proof' in mathematics. The problem to be faced is, of course, that it is not possible to actually *exhibit* an infinite number of primes; you must somehow prove that they are infinite in number without actually producing them all.

The idea is to show that *if you were* to start listing all the primes, your list *would* continue for ever. To do this, let us agree to denote the primes in your list by the symbols $p_1, p_2, p_3$, etc. So $p_1$ is the first prime, namely 2, $p_2$ is the next, namely 3, $p_3$ the third, namely 5, and so on. This use of number subscripts to denote the members of a list is very common in mathematics: at a glance one knows that $p_{88}$ denotes the 88th prime (whatever that is) in the list. What we want to show is that the list $p_1, p_2, p_3, \ldots$ (where the dots mean 'continue the list as far as possible') continues indefinitely. Suppose then that we have (hypothetically) listed all the primes up to the $N$th, where $N$ is some large, but unspecified stage, obtaining the list $p_1, p_2, p_3, \ldots, p_{N-1}, p_N$. How can we be sure that the list does not stop at this point? This is where you need to be clever. The trick is to look at the number formed by multiplying together all the primes in your list and then adding 1; that is, look at

$$M = p_1 \times p_2 \times p_3 \times \ldots \times p_{N-1} \times p_N + 1$$

This number will likely be astronomically large but no matter, we need to form it only 'in theory'. The number $M$ is (much) bigger than $p_N$. So if $M$ is prime we know that the list of primes will not stop at $p_N$. (It may well be that $M$ is not the *next* prime after $p_N$, but that is not important; once we know that $p_N$ is *not the last* prime our task is complete.) What if $M$ is not a prime? Then, by the Fundamental Theorem of Arithmetic $M$ will be a product of primes. Now, any prime which occurs in this product will divide exactly into $M$. But if any of the primes $p_1, p_2, \ldots, p_N$ is divided into $M$ there is obviously a remainder of 1. (This is why we added that 1 when we formed $M$.) So $M$ is a product of primes which do not occur in the list $p_1, p_2, \ldots, p_N$. So in this case also we conclude that there must be primes beyond $p_N$. The inescapable conclusion now is that there are indeed an infinite number of primes.

The curiosity of mankind being what it is, it is not surprising that there has been considerable interest in discovering 'largest known' primes, a curiosity fuelled by the availability of ever greater computer power. But computing power alone is not enough to win at the 'largest prime number in the world' game; you need some mathematical knowledge too. The problem is how do you test if a given number is prime or not. Naively, to see if $N$ is prime, you look at each of the numbers 2, 3, 4, 5, . . ., $N-1$ in turn and see if any of them divides (exactly) into $N$; if one does, then $N$ is not prime, if none does then $N$ is prime. This can be speeded up somewhat by observing that if $N$ is not prime, it will be divisible by some number which is not greater than the square root of $N$, so you need to look only for possible divisors up to the square root of $N$. To further simplify matters, once you have checked whether 2 is a divisor, if it is not then there is no need to look at any other even numbers. Likewise, if 3 is not a divisor, any multiples of 3 may be eliminated from the search. Taken to its logical conclusion, of course, it is really only necessary to look for possible divisors among the primes themselves; but this begs the question, since what we are after is a method to test for primality, and this method should not depend upon other primality tests (or even the same test) along the way.

That last remark needs a little amplification. For relatively small numbers, looking for possible divisors is quite feasible; either by storing a table of primes or else by looking at, say, 2 plus all odd numbers. (And in a sense the former approach does depend upon a previously run primality test.) For instance, there are only 168 primes less than 1000, and by using these as trial divisors it will be possible to test the primality of any number less than 1,000,000. But if you want to use the same method for testing primality of numbers of the order of, say, a million million million (that is, numbers with around 18 digits) you would need to have available over 5 million primes, or else be prepared to carry out half a billion trial divisions. And 18 digit numbers are pretty small fry: for instance, some cryptographic systems in use today (see chapter 13) involve prime numbers with a hundred digits. In fact, trial division as a method of testing primality rapidly becomes infeasible as the size of the number increases. For instance, the fastest computers currently in use can perform something like 200 billion arithmetic operations per second. Using such a machine, to test for primality by trial division would require 2 hours of computer time for a 20 digit number, 100 billion years for a 50 digit number, and for a 100 digit number a staggering

million million million million million million

years. Fortunately for prime number hunters, however, there are alternative methods for testing primality (see chapter 8). Using one of the most efficient of these, a test developed by the mathematicians Adleman, Rumely, Cohen and Lenstra, and named after them, the timings corresponding to the above are 10 seconds for 20 digits, 15 seconds for 50 digits, and 40 seconds for 100 digits.

But even clever tests like the Adleman–Rumely–Cohen–Lenstra are no good for finding record primes. Since the 1950s, the largest known primes have all had in excess of 1000 digits (see the table on page 102), and for a 1000 digit number this test would take about one week. Remember, in order to *find* a new prime you have to run the test on lots of numbers, one after the other, until you find one that is prime. ('Most' numbers are not prime, of course. A half of them are even for a start, and one in three of the odd numbers is a multiple of 3.) And when you consider that the current record holder is a prime number with nearly forty thousand digits, it is clear that something else is going on.

Record primes are nowadays all numbers of the form

$$2^N - 1$$

Numbers of this form are called *Mersenne numbers* after a seventeenth century French monk of that name, who made some (amazingly accurate) conjectures about the primality of these numbers. In his book *Cogitata Physica Mathematica* (1644), he claimed that the number $2^N - 1$ is prime for values of $N$ equal to 2, 3, 5, 7, 13, 19, 31, 67, 127, 257, and fails to be prime for all other values of $N$ less than 257. It was not until 1947, some 300 years later, that desk calculators were used to discover that Mersenne had made a couple of errors: $N$ equal to 67 and 257 do not yield Mersenne primes, and the values 61, 89, 107 do. The astonishing degree of accuracy of Mersenne's claim can be appreciated when you gain some idea of the size of Mersenne numbers.

To try to appreciate the size of Mersenne numbers, a good example to look at is the number $2^{64}$, just one more than the Mersenne number $2^{64} - 1$. This can be 'visualised' as follows. Take an ordinary chessboard, and number the squares from 1 to 64. (It does not matter whether you number the squares row by row or column by column.) On square 1, place two 10p coins. On square 2 put four 10p coins. Put eight on square 3, sixteen on square 4, and so on. Each time, you put twice as many coins on the square as you did on the previous one. Now, a single 10p coin is 2 mm thick. On square number 64, you will have a pile of exactly $2^{64}$ coins. How high do you think this pile will be? 1 metre? 50 metres? 100 metres? A kilometre even? Wait for it. The pile will be just under 37 million million kilometres high. So your pile will stretch out beyond the Moon (a mere 400,000 kilometres away) and the Sun (150 million kilometres from Earth), and in fact will reach (just) the nearest star, Proxima Centauri, some 4 light years from Earth. Written out fully, the number $2^{64}$ looks like

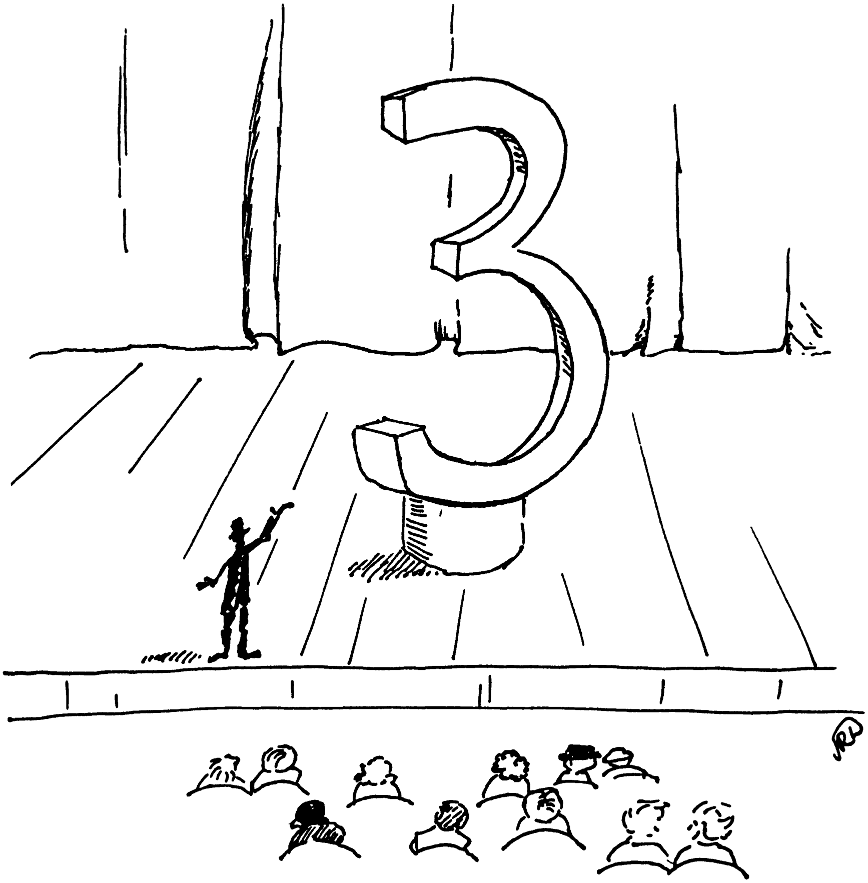
18,446,744,073,709,551,616

Try to imagine now the number $2^{19937} - 1$. In 1971, IBM's Bryant Tuckermann used an IBM 360-91 computer to show that this number is prime. This broke the previous record, $2^{11213} - 1$, which had stood since its discovery in 1963 using the old ILLIAC–II computer. Tuckermann's number has some 6002 digits, and its discovery began a hunt for record primes using very powerful computers which has continued to this day. Record prime hunters restrict their search to Mersenne numbers because there is a very clever method for testing primality of Mersenne numbers invented by Lucas and improved by Lehmer, and named after them as the *Lucas–Lehmer test* (see page 51). This test capitalises on the fact that the size of the number $2^N - 1$ increases rapidly as $N$ increases by small amounts. The computation time for the test on $2^N - 1$ depends upon the size of $N$ rather than on the (astronomical) size of the number $2^N - 1$ itself.

Two 15 year old high school students of Hayward, California, upon reading of Tuckermann's discovery, decided they would try to better it. From 1975 until 1978 they spent their time finding out how to go about discovering a new record prime, and writing a computer program that would do the job. After some 350 hours of computer time at the computer centre of the University of California at Hayward, the two students, now 18 years old, found their record prime: the 6533 digit number $2^{21701} - 1$. Young Laura Nickel and Curt Noll and 'their' CDC–CYBER–174 computer became instant celebrities. Their discovery was front page news across the United States and was reported on nationwide television. Now everyone knew about primes and the incredible computing power of modern computers, even in the hands of a couple of teenagers.

In 1979, Noll bettered the record with the 6987 digit number $2^{23209} - 1$, but only just got there before David Slowinski, a young programmer for Cray Research, who brought the immensely powerful CRAY-1 computer into the game. During the period from 1976 to 1982, this was arguably the most powerful computer in the world. (It was certainly the one with the fastest 'clock time', the time taken for the computer to change from one internal 'state' to another: the CRAY-1 does this in just 12.5 billionths of a second.) Slowinski and his CRAY-1 were just a couple of weeks too late in their discovery of the prime $2^{23209} - 1$, but Noll's record was not to last long. A short while later, Slowinski, aided by Harry Nelson, discovered the 13,395 digit monster prime $2^{44497} - 1$. In September 1982, Slowinski and the CRAY-1 took the record up to $2^{86243} - 1$, a number with 25,962 digits. And then the current record, a prime with 39,751 digits, was found in September 1983 using a CRAY-XMP computer, an upgraded CRAY-1 machine. This number, $2^{132049} - 1$, begins with the sequence 51274 and ends with 61311. The Lucas-Lehmer test took just over one hour to show that this number was prime. The search for it lasted six months, during which time two Cray computers were used, non-stop.

Why bother? To some extent this is like asking why people climb mountains. But for the computer manufacturer there are certainly two tangible rewards to be gained. Firstly, running a primality searching program which deals with numbers of the size of record primes is a good way of testing the computer hardware and software; and Slowinski made use of computers undergoing 'factory testing', so in a sense the computer time used was all 'free'. And secondly, there is, the world being as it is, a great deal of publicity to be had for the computer firm which makes the machine that finds the prime. Record primes have little interest for the professional mathematician, but they certainly have a habit of hitting the newspapers and TV screens.



LADIES + GENTLEMEN, THE BIGGEST PRIME NUMBER IN THE WORLD!

## *Twin Primes*

The distribution of the prime numbers among all the whole numbers appears to follow no particular pattern. There are arbitrarily long sequences of numbers which contain no primes at all, while at the other end of the spectrum there occur pairs of successive odd numbers both of which are prime – for example, 3 and 5, or 11 and 13, or 1,000,000,000,061 and 1,000,000,000,063. Such pairs of successive odd numbers which are prime are called *twin primes*.

Computer searches have shown that there are 152,892 pairs of twin primes less than 30,000,000. Twin primes appear to be less frequent as the size of the numbers increases, but it is not known if there are infinitely many twin prime pairs or not. The 'Twin Prime Conjecture' asserts that there *are* infinitely many.

If you can solve the Twin Prime Problem, not only will you be famous overnight, but you will also be better off financially. Worldwide Computer Services in Wayne, New Jersey, USA, has offered a prize of $25,000 to the first person to settle the Twin Prime Conjecture one way or the other.

******

On a rather more practical level for most mortals, rather than trying to prove the Twin Prime Conjecture you might like to try hunting for some large primes yourself. The method used to find world record primes is explained on page 51, though you may well prefer to play a more modest game, like trying to find the largest prime that fits in one computer word, or two, or three, etc. The only other ingredient you need – besides your micro, of course – is the knowhow to write some routines to handle large numbers in the computer. The easiest way to do this is just to take the standard rules that you learnt in school to perform arithmetic on numbers with more than one digit each, using 'column position' to denote whether the digit represents a unit, tens, hundreds, or whatever. For the computer, the analogue of a single digit would be an entire computer word, though to keep things reasonably straightforward it might be better to work instead with numbers which occupy at most half a word, so as to avoid any risk of overflow during multiplication. Most people like to write their own routines for performing 'multiple precision arithmetic', as arithmetic with very large numbers is called, but in case you need it, chapter 16 at the very end of the book should give you some help. (You will also find there a few ideas for speeding up some multiple precision arithmetic routines. The bright backroom guys have been at work in this area as well!)

### *The Sum Total*

1. Using the digits from 1 to 9 inclusive, each once, you can write down a single fraction which is equal to 1/2. Namely

   7293/14586.

   Now do the same thing but with a fraction equal to 1/3.
2. If you are still feeling smug after doing question 1, do the same thing again to get answers equal to each of 1/4, 1/5, 1/6, 1/7, 1/8, 1/9. Yes, they can all be achieved.
3. Arrange the digits from 0 to 9 into two fractions whose sum is 1.
4. Which two digit number am I talking about when I say that if you triple it and then add the two digits of the original number the result is the original number with the digits reversed?
5. Two numbers consist of the same two digits reversed. The smaller number is one less than one-half the larger number. What are the two numbers?

Answers to all of these teasers can be found at the back of the book. They should be looked up only when insanity is imminent.

# 2 *Pi and chips*

As every schoolchild knows, to calculate the circumference of a circle of diameter $d$ you multiply $d$ by the number $\pi$ ('pi'). The value of $\pi$ is commonly taken to be 22/7, but this is only an approximate value. As a decimal 22/7 is

3.142859 142859 142859 . . .

where the pattern 142859 repeats endlessly. The decimal expression for $\pi$ on the other hand continues indefinitely without any regular pattern setting in (to describe this fact, mathematicians say that $\pi$ is *irrational*), commencing with the sequence

3.14159 26535 89793 23846 . . .

So 22/7 is accurate to only two decimal places.

Since it requires an infinite number of decimal places to give the value of the number we call $\pi$ with total accuracy, how is the number specified in the first place? Certainly not by giving its value, of course! In fact, $\pi$ is *defined* to be the ratio of the circumference of any circle to its diameter. Besides implying that the above quoted formula for the circumference of a circle does not have any real content, this definition pre-supposes a rather amazing fact: namely that no matter what size circle you take, be it a few centimetres in diameter or many kilometres across, the answer you get when you divide the circumference by the diameter is always the same.

Supposing you wanted to calculate the value of $\pi$. How could you proceed? You could draw a circle of diameter, say, 1 metre. Then its circumference would be $\pi$ metres. But how can you determine the length of the circumference? Measuring the length of a circumference is so difficult that one usually resorts to calculating it using the formula stated above: which, as we have seen, does not help us if the aim is to calculate $\pi$ in the first place. The idea is to approximate the circle by means of a polygon with a sufficiently large number of sides, as shown in figure 1.

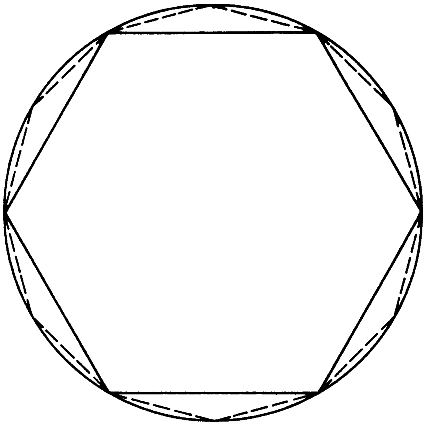


*Figure 1.* To calculate an approximate value for the circumference of a circle, evaluate the total length of the edges of a polygon drawn inside the circle as shown. The more sides the polygon has, the better this approximation will be

Measuring the straight sides of such a polygon is an easy matter and, if the polygon has enough sides, this measurement differs from the actual circumference by only a small amount. The more sides the polygon has, the better the approximation. In fact there is no need to restrict this approach to actual, physical measurements. Using elementary ideas of geometry, if the polygon is a regular one (that is, if all its sides are the same length), the length of each side can be calculated from a knowledge of the number of sides. Using this idea, in the third century B.C. Archimedes calculated that $\pi$ was approximately equal to 22/7. And by A.D. 150 the value 3.1416 was known. These values are considerably more accurate than the value 3 which is implied by two passages in the Bible, I Kings 7.23 and II Chronicles 4.2. To quote the former

> *And he made a molten sea, ten cubits from the one brim to the other; it was round all about, and his height was five cubits: and a line of thirty cubits did encompass it round about.*

The second passage is similar.

The fact that the decimal expression for $\pi$ continues indefinitely without settling down to any repetitive behaviour has been known for certain since 1882 when Lindemann succeeded in proving this fact. Indeed, Lindemann proved rather more, namely that $\pi$ is not the root of any polynomial with integer coefficients (in formal terms, $\pi$ is *transcendental*), a result which implies that the ancient problem of squaring the circle using ruler and compasses alone is impossible. (Not that the known impossibility of this task since 1882 has prevented numerous amateur mathematicians from continuing to try to do just that, even to this day.) If anything, this knowledge that $\pi$ is transcendental has spurred on attempts by mathematicians to calculate the decimal expression for $\pi$ to ever greater degrees of accuracy.

In 1596, the German mathematician Ludolph van Ceulen calculated $\pi$ to 35 places of decimals, and in accordance with his wishes his 35 places were inscribed on his tombstone when his death at the age of 70 finally put a stop to his calculations. (German mathematicians still sometimes refer to $\pi$ as the *Ludolphian number*, though the ever-increasing use of English in mathematics over the recent few decades appears to be killing off this somewhat touching custom.)

Computation of $\pi$ became easier with the invention of the calculus in the seventeenth century, which brought with it various infinite expressions for $\pi$ (see page 53 for a brief discussion of infinite sums and what they mean). Leibnitz obtained the formula

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \ldots$$

where the sum continues for ever in the manner indicated, with the denominators going up through all the odd numbers and the sign altering at each stage. Because the terms in this sum become smaller and smaller as you go out along it, by calculating the sum of, say, the first fifty terms you get a moderately acceptable approximation to $\pi$. (But since there are much better methods, it is not worth dwelling on this one here.) At about the same time, Wallis derived the formula

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{7} \cdot \frac{8}{9} \cdot \ldots$$

which is an infinite product. The formula

$$\frac{\pi}{4} = 4\left(\frac{1}{5} - \frac{1}{3 \times 5} + \frac{1}{5 \times 5} - \frac{1}{7 \times 5} + \ldots\right) - \left(\frac{1}{239} - \frac{1}{3 \times 239} + \frac{1}{5 \times 239} - \frac{1}{7 \times 239} + \ldots\right)$$

was obtained by Machin at the beginning of the eighteenth century, and gives very accurate values for $\pi$ using only a few terms. (It is clear that the terms in this formula grow small very rapidly indeed.)

In 1699, Abraham Sharp calculated $\pi$ to 71 decimal places. In 1824, a chap called Dase, a lightning calculator employed by Gauss, worked out 200 places. In 1854, Richter got to 500 places. During the nineteenth century, a gloriously eccentric English mathematician called William Shanks devoted 20 years of his life in calculating $\pi$ to 707 decimal places; he published his result in the *Proceedings of the Royal Society* in 1873–4. Unfortunately, in 1945, using desk calculators, a mistake was found in the 527th and subsequent places of Shanks' result, but of course Shanks was by then long past caring.

In recent times, computers have made the calculation of $\pi$ much easier, of course. In 1973, Guilloud and Bouyer in France published as a book the first one million places of $\pi$. For the record, the book ends with the sequence

. . .5779458151

In 1981, after 137 hours of computation on a FACOM M200 computer, Kazunori Miyoshi of the University of Tsukuba, Japan, obtained two million places, and had to decide what to do with the 800 pages of print-out this required. In 1983, Yoshiaki Tamura and Yasumasa Kanada of the University of Tokyo Computer Centre calculated $\pi$ to 8 million decimal places. The HITAC M–280H computer they used was so powerful that the calculation took a mere 7 hours. Then, to be absolutely sure of their record, they continued up to 16 million places, but it turned out that the result could be relied upon only up to place 10,013,395.

The problem with calculations of $\pi$ to enormous numbers of places is that, as the calculation proceeds, small errors can accumulate, which eventually lead to an incorrect digit. To guard against this possibility, Tamura and Kanada made a second calculation of $\pi$ using another program, this time running on a new Japanese 'supercomputer', a Hitachi S–810 model 20 computer. The calculation took 24 hours, after which the two results were compared. They agreed up to place 10,013,395, thereby guaranteeing the result up to that stage.

"Why bother?" you may ask. Just as with the search for large primes (see chapter 1), pure curiosity accounts for some of the motivation. There is also the fact that such a prolonged calculation provides a good method for testing new computer hardware and software. To say nothing of the publicity the computer manufacturer gets when the new record is announced.

It is also just possible (though extremely unlikely) that by examining (using the computer) the decimal expansion of $\pi$ some pattern may be discerned which could lead to new mathematical discoveries being made. The point is that, because the decimal expression for $\pi$ is produced by a formula, the sequence of digits in this expression cannot constitute a truly *random* sequence, but so far as it has been investigated the sequence does behave like a random sequence, passing with flying colours all the tests for 'randomness' which statisticians have devised.

One property that a random sequence of digits will possess is that any given finite sequence of digits will occur somewhere in the sequence. For instance, in a random sequence, the finite sequence 123456789 will occur somewhere. Tamura and Kanada have found that this does not happen in the first 10 million places, though the sequence 23456789 does occur, starting at place 995,998. The longest sequence of consecutive zeros they have found has length seven and starts at place 3,794,572. Also, starting at place 1,259,351 you find the sequence 314159 which commences the expression for $\pi$.

All of which means that we have come a long way from the Babylonian value of 3.125 obtained over 4000 years ago. Though even that value is much better than that which, in 1897, was declared to be used in the State of Indiana, USA: in that year the General Assembly enacted a bill to the effect that $\pi$ is equal to 4. I have no idea how long this bill remained on the statutes, though I can imagine that the local wine merchants lobbied long and hard for its retention.

### *A Head for Figures?*

The calculation of $\pi$ to many decimal places does provide the rest of mankind with one rather dubious benefit. People can spend their time memorising the expression to record lengths. The current record holder is Rajan Srinivasen Mahadevan of India, who, in 1981, correctly recited 31,811 places, the recitation taking an astonishingly fast 3 hours and 49 minutes. The current UK record holder is Creighton Carvello of Redcar, who memorised 20,013 places in 1980. Besides having a good memory, Carvello was presumably in pretty good physical shape as well, since it took him over 9 hours to recite the thing.

******

### *Pi in the Sky?*

References in the Bible (I Kings 7.23 and II Chronicles 4.2) to $\pi$ being equal to 3 have led a group of Kansas academics to form The Institute for Pi Research, whose main aim is to propagate the use of the value of 3 for $\pi$. As the Institute's founder, Samuel Dicks, professor of medieval history says, "If a pi of 3 is good enough for the Bible, it is good enough for modern man."

One of the Institute's aims is to get state schools to give $\pi = 3$ equal time with the more conventional value. Coupled with Dicks' remark that the Pi Institute deserves to be taken as seriously as the Creationists, this leads one to suspect the real aim of all of this. But whatever they are really after, they may have some friends in high places. The Institute sent a letter to US President Reagan asking for support, and though they did not receive a reply they were encouraged to hear him say in a speech shortly afterwards that "The pi(e) isn't as big as we think."

# 3 Formulas for primes

The distribution of prime numbers among all whole numbers seems to be so erratic that no simple formula could exist which would produce as its values all, and only, the primes. If by 'formula' you mean here 'polynomial formula', then this is true. For instance, to take the simplest case of a polynomial formula, namely a linear formula of the form

$$f(n) = An + B$$

where $A$ and $B$ are constants, for infinitely many values of $n$, $f(n)$ will fail to be a prime. This is easy to check for yourself. Much more difficult to establish is a famous result of Dirichlet, a nineteenth century mathematician, which says that $f(n)$ will, however, be prime for infinitely many values of $n$.

A natural question to ask is what is the longest sequence of prime numbers which can be produced by a formula of the form

$$f(n) = An + B$$

for values of $n$ equal to 0, 1, 2, 3, etc. in turn. The current record is held by Paul Pritchard, a computer programmer at Cornell University, USA, who used a DEC VAX 11 supermini computer to find a formula which produces 18 primes in a row in 1983.

The task facing Pritchard was not particularly hard. The mathematics required to write a program which looks for formulas that produce 'long' sequences of primes is quite straightforward and well known. What you need is lots of time on the computer. Pritchard obtained his computer time in a particularly efficient manner. Most mainframe computers and superminis like VAX are so efficient (and so expensive) that they are in constant use, 24 hours a day, throughout the year. To make maximum usage of the machine possible, it is generally equipped with a number of separate terminals, where different users can access it at the same time. A sophisticated control program called an operating system shares out the computer's time between the different users, both for input (from a keyboard or magnetic tape or disk) and output (to a screen or to tape or disk), as well as for actual computation. Modern computer speeds are such that fifty or so users can be accessing the machine at the same time without any one of them being aware that they are not alone on the machine. In fact, even a 'heavily used' computer will still spend most of the time sitting 'idle', waiting for someone to instruct it what to do next. (Remember, today's computers are capable of performing millions of instructions per second.)

What Pritchard did was to make use of this 'idle' time in making his search for a prime-producing formula. He instructed the computer to work on his problem whenever there was nothing else to do, and drop it when something cropped up. With this approach, it turned out that the computer was able to devote around 10 hours to the problem every day. Within a month of starting his search, Pritchard got what he wanted, a formula which gives 18 primes, breaking the old record by 1. The formula he (or rather his computer) found is

$$f(n) = 9{,}922{,}782{,}870\,n + 107{,}928{,}278{,}317$$

This formula gives a prime value for $f(n)$ for $n$ equal to 0–17.

A related problem is to find formulas whose successive values are *consecutive* primes. The record to date is a sequence of 6 consecutive primes, produced by the formula

$$f(n) = 30n + 121{,}174{,}811$$

for $n$ equal to 0–5.

When you come to look at quadratic formulas, the result is a little better. The record holder is the formula

$$f(n) = n^2 + n + 41$$

discovered by the great eighteenth century mathematician Leonhard Euler. The values of this formula are prime for all values of $n$ from 0–39; that is, an unbroken sequence of 40 primes. For $n = 40$, you get the value $f(40) = 41^2$, which is not prime, but even then you continue to get lots of primes from this formula. Indeed, of the first 2398 values, exactly half are prime, while of the first 10 million values the proportion of primes is 0.475. . ., not far short of half.

Euler's quadratic formula seems to be unique. No other is known which produces anything like as many consecutive primes. Using a

VAX 11 supermini computer some time ago, I examined all quadratic polynomials of the form

$$f(n) = An^2 + Bn + C$$

for every possible combination of values of the constants $A$, $B$, $C$ from 0 to 1000, and then with values of $A$ between 1 and 100 and $B$, $C$ between 0 and 10,000. So, in all I (or rather my VAX) looked at well in excess of 10 billion formulas. None was found which could better Euler's formula, produced 300 years ago. Indeed, nearly all failed quite miserably. The quite remarkable nature of the Euler formula has led some mathematicians to think that there may be some deep and as yet unknown reason for its behaviour (see chapter 10).

Though it is not possible for a polynomial (see page 21) formula to generate all the primes (and no other numbers), there are various relatively simple formulas which do the trick. The nicest one that I know of is the following. To make the formula easier to understand, I shall split it up into two parts. First comes the formula

$$h(m, n) = m \times (n + 1) - (n! + 1)$$

For any two (whole number) values for $m$ and $n$, the value of $h(m, n)$ is readily calculated, provided that you understand the meaning of the mathematician's notation

$$n!$$

(This is read as '*n factorial*'.) This is shorthand for the product of all the whole numbers from 1 to $n$ inclusive. Thus the first few factorial values are

$$2! = 2 \times 1 = 2$$
$$3! = 3 \times 2 \times 1 = 6$$
$$4! = 4 \times 3 \times 2 \times 1 = 24$$
$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

Try working out the values 6! to 10! yourself. This involves less effort than might at first be supposed, since each successive factorial value can be obtained from the previous one by a single multiplication. One thing that will become immediately clear when you do this is that the factorial numbers grow large very rapidly. (10! is already well into the millions.)

Having described the formula $h(m, n)$, the prime-generating formula that we are aiming for is

$$f(m, n) = \tfrac{1}{2}(n - 1)[\mathrm{ABS}(h(m, n)^2 - 1) - (h(m, n)^2 - 1)] + 2$$

(The formula ABS($k$) which is used here is the *absolute value* function, which simply discounts any minus sign that $k$ may have. So, for example, ABS(3) = 3 and ABS(−5) = 5.) For any values of $m$ and $n$, the value of $f(m, n)$ is prime, and all primes are values of $f$ for some numbers $m$ and $n$. But a few moments' experimentation with this formula indicate that it is not a very efficient generator of primes. For most values of $m$ and $n$ you get the value $f(m, n) = 2$; in fact $f(m, n) = 2$ for infinitely many values of $m$ and $n$. But occasionally $f(m, n)$ takes a value other than 2, and each time this occurs a new prime number is produced. In fact, the odd primes are each produced exactly once by the formula.

For $m = 1, n = 2$ you get the value 3, while for $m = 5, n = 4$ you get 5. The next two odd prime values are $7 = f(103, 6)$ and $11 = f(329891, 10)$, which gives some indication as to just how 'rare' is the production of an odd prime by this formula. This rarity is caused by the rapid growth of the factorial function in the formula $h(m, n)$. The only time when $f(m, n)$ produces a result other than 2 is when $h(m, n) = 0$, when you get $f(m, n) = n + 1$. To get $h(m, n) = 0$ you must have $m \times (n + 1) = n! + 1$, so if $n$ is reasonably large, $m$ has to be enormous.

The mathematical fact which lies behind the above formula is known as *Wilson's Theorem*. John Wilson was a minor eighteenth century English mathematician who noted that if $n$ is a prime number, then $n$ divides exactly into the number $(n - 1)! + 1$. In fact, Wilson was not up to providing a mathematical proof of this fact, nor indeed was his teacher, the famous mathematician Edward Waring; but in 1771 Lagrange supplied such a proof. So Wilson was lucky in that, simply by guessing the result on the basis of numerical evidence, he managed to achieve some sort of immortality. At any rate, not only did Lagrange prove Wilson's Theorem, he showed also that the converse is true: any number $n$ for which $n$ divides into $(n - 1)! + 1$ must be prime. Thus, perhaps surprisingly, the property that a number $n$ divides into $(n - 1)! + 1$ exactly characterises the primes. Using this fact, it is an easy exercise to verify that the formula $f(m, n)$ given above does indeed generate each odd prime exactly once.

## *A Prime Candidate*

Though there cannot be a polynomial formula which generates all and only the prime numbers, if you allow yourself the use of more than one variable and agree to the possibility of the formula producing negative, non-prime values from time to time, then you can have a prime-producing polynomial formula. The following formula involves 26 variables (an amazing stroke of luck, since there are just 26 letters of the alphabet which can be used to label these variables) and has degree 25. When (non-negative) whole numbers are substituted for these 26 variables, the positive values produced by the formula are precisely the prime numbers. The polynomial also produces negative values, which need not be prime.

$$\begin{aligned}
&(k+2)\{1-[wz+h+j-q]^2-[(gk+2g+k+1).\\
&(h+j)+h-z]^2-[2n+p+q+z-e]^2\\
&-[16(k+1)^3.(k+2).(n+1)^2+1-f^2]^2\\
&-[e^3.(e+2)(a+1)^2+1-\sigma^2]^2-[(a^2-1)y^2\\
&+1-x^2]^2-[16r^2y^4(a^2-1)+1-u^2]^2\\
&-[((a+u^2(u^2-a))^2-1).(n+4dy)^2+1\\
&-(x+cu)^2]^2-[n+l+v-y]^2-[(a^2-1)l^2+1-m^2]^2\\
&-[ai+k+1-l-i]^2-[p+l(a-n-1)\\
&+b(2an+2a-n^2-2n-2)-m]^2-[q+y(a-p-1)\\
&+s(2ap+2a-p^2-2p-2)-x]^2-[z+pl(a-p)\\
&+t(2ap-p^2-1)-pm]^2\}
\end{aligned}$$

(There is no paradox caused by the fact that the formula above seems to have a factor of $(k + 2)$. The formula works by the remaining factor producing only a positive result of 1, this occurring in precisely those cases when $k + 2$ is prime.)

The formula was found by James Jones, Daihachiro Sato, Hideo Wada and Douglas Wiens in 1977, after Martin Davies, Yuri Matijasevic, Hilary Putnam and Julia Robinson had proved that such a formula had to exist.

## *Pseudoprimes*

Wilson's Theorem shows that it is possible to characterise prime numbers other than by means of the definition. Over 25 centuries ago, Chinese mathematicians thought they had found an alternative characterisation of prime numbers. They claimed that a number $n$ will be prime if, and only if, $n$ divides exactly into $2^n - 2$. In the seventeenth century, the great French mathematician Pierre De Fermat did prove that, if $n$ is prime, then $n$ divides into $2^n - 2$, but there are non-prime numbers with this property too, so it does not exactly characterise the primes. The first non-prime with the property is $341 = 11 \times 31$. There are only two others that are less than 1000, which perhaps explains why the Chinese, equipped with only the abacus, fell into the trap of thinking that they had found a universally true law of arithmetic.

A non-prime number $n$ which divides into $2^n - 2$ is called a *pseudoprime*. You might like to try to find all 22 pseudoprimes that are less than 10,000.

There are some examples of *even* pseudoprimes. You could try to find one of them as well. Though fairly big, the first one of these should be accessible to the average home micro.

Modern high-speed primality tests work by a refinement of the above Chinese property which avoids the difficulties caused by the existence of pseudoprimes.

# 4 *The kilderkin approach through a silicon gate*

What is the difference between a modern electronic computer and a thirteenth century English wine merchant? The answer is "Not as great as you might think." The major clue lies in the system of measurement used in the wine and brewing trade in England from the thirteenth century onwards, parts of which are still in use

2 gills = 1 chopin
2 chopins = 1 pint
2 pints = 1 quart
2 quarts = 1 pottle
2 pottles = 1 gallon
2 gallons = 1 peck
2 pecks = 1 demibushel

2 demibushels = 1 bushel or firkin
2 firkins = 1 kilderkin
2 kilderkins = 1 barrel
2 barrels = 1 hogshead
2 hogsheads = 1 pipe
2 pipes = 1 tun

As you can see, thirteenth century wine merchants in England measured their wares using a system of counting based on the number 2, what we now call the binary system of arithmetic. Leaving aside the wonderfully evocative vocabulary of the above system, this means that they performed their arithmetic in the same way that a modern computer does.

We are so used to computers nowadays that it seems obvious that arithmetic should be performed in a binary fashion, this being the most natural form for a computer, which is, ultimately, a 'two-state' machine (the current in a circuit may be either on or off, an electrical 'gate' may be either open or closed, etc.). But this was not always the case. When the first American high-speed (as they were then called) electric computers were developed in the early 1940s, they used decimal arithmetic, just like their inventors. But in 1946, the mathematician John von Neumann (essentially the inventor of the 'stored program' computer that we use today) suggested that it would be better to use the binary system of arithmetic, since which time binary computers have been the norm. (Not that this was the first time that calculating machines made use of the binary system. Some French machines developed during the early 1930s used binary arithmetic, as did some early electric computers designed in the United States – by John Atanasoff and by George Stibitz – and in Germany – by Konrad Zuse.)

There is, of course, nothing special about the decimal number system that we use every day. Certainly it was convenient in the days when people performed calculations using their fingers. Assuming a full complement of these, it is essential that there is a 'carry' when we get to ten. The number at which a 'carry' occurs in any number system is called the 'base' of that system. In base 10 arithmetic (decimal arithmetic), 10 entries in the units column are replaced by 1 entry in the 10s column, 10 entries in the 10s column by 1 entry in the 100s column, and so on. This means that we require ten 'digits' in order to represent numbers, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, all other numbers being composed of a string (or 'word' if you like) made up from these digits. Computers (and electronic calculators) use the binary system to perform their arithmetic. Here there are only two digits (known as 'bits', short for 'binary-digits'), 0 and 1. In binary arithmetic there is a 'carry' whenever a multiple of 2 occurs. So, counting from one to ten in binary looks like

1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010

Arithmetic in binary (addition, multiplication, etc.) is performed just as in the decimal arithmetic that we learn in primary school, except that we 'carry' multiples of 2 into the next column rather than multiples of 10. (So instead of having a units column, a tens column, a hundreds column, and so on, we have a units column, a twos column, a fours column, an eights column, a sixteens column, and so on.) It is the fact that in binary notation all numbers can be expressed using just two digits, 0 and 1, that makes the binary system particularly suited to electronic computers. As I mentioned earlier, the ultimate construction element of a computer is an electrical switch that is either on or off (1 or 0) – a 'gate'.

Of course, we do not use binary notation when we communicate with a computer or a calculator. We feed numbers into the machine in the usual decimal form, and the answer comes out in this form as well. But the computer/calculator immediately converts the number into binary form before commencing any arithmetic and converts back into decimal form to give us the answer. What should be emphasised is that it is just a matter of notation (or language, if you like) that is involved here. The actual numbers are the same. 111 in binary means the same as 7 in decimal notation, just as *das Auto* in German means the same as *the car* in English.

All of this is a good excuse for bringing in the following teaser, one which can be used to demonstrate the absurdity of many of the questions beloved by testers of IQ in children. Fill in the next two members of the following sequence

10, 11, 12, 13, 14, 15, 16, 17, 20, 22, 24, 31, 100, –, –

The answer is given at the back of the book.

### *Palindromic Numbers*

Most people are familiar with linguistic palindromes, sentences which read the same backwards as forwards, such as Adam's greeting to Eve upon meeting in the Garden of Eden: "Madam, I'm Adam." Palindromic numbers are just the numerical equivalent of these, numbers which read the same both ways, such as 12321 or 1890981. In themselves, palindromic numbers are not at all interesting, of course, since they can be made up so easily. They become more interesting when you ask for palindromic numbers of a certain kind. For instance, are there perfect squares which are palindromic?

Yes there are. For instance, 11 X 11 = 121, 26 X 26 = 676, and 264 X 264 = 69696. In fact, palindromic squares are fairly common. Try writing a program to list palindromic squares. You will soon notice a rather curious fact. The palindromic numbers all seem to have an odd number of digits. Not until you reach the numbers 836 X 836 = 698896 will you see a palindromic square with an even number of digits. The next two are

$$798644^2 = 637832238736$$

and

$$64030648^2 = 4099923883299904$$

Early in 1984 I wrote about palindromic squares with an even number of digits in *The Guardian*. At the time the only example I knew of was the first one quoted above. Numerous readers discovered the other two given, but only one person managed to find a fourth example.

Graham Lyons of Romford in Essex ran his IBM Personal Computer over an entire weekend to discover the 22 digit palindromic square

$$83163115486^2 = 6916103777337773016196$$

As far as I know, this remains the record, so the field is all yours. I should point out that it is advisable to spend a bit of time looking at the problem mathematically before you set your computer off on its hunt, as there are a lot of numbers that you will have to look at! One hint which may be helpful is that any palindrome with an even number of digits must be divisible by 11. Proving this odd little fact is in itself a pleasant exercise.

Good hunting! And don't forget to let me know of any successes.

# *5 Colouring by numbers*

Early in 1984, the Fredkin Foundation of Boston, Massachussetts, USA, offered a prize of $100,000 to the first person to write a computer program which subsequently makes a genuine mathematical discovery. All entries are to be examined by a twelve member committee of experts headed by Woodrow Bledsoe, Professor of Computer Science at the University of Texas at Austin. Bledsoe is a leading figure in that area of computer science which deals with attempts to program computers to prove mathematical theorems.

The task facing the would-be winner of the Fredkin prize is by no means an easy one. According to Bledsoe, "The prize will be awarded only for a mathematical work of distinction in which some of the pivotal ideas have been found automatically by a computer program in which they are not initially implicit." So there you have it. The computer must somehow make part of the discovery itself, and not be just the workhorse of a clever mathematician.

Looking back over the use of computers in mathematics over the past thirty years or so, I can see nothing that would come remotely close to winning the prize. Computers have certainly played an important role in several mathematical discoveries, but on each occasion it is the human mind that has provided all the essential ideas. The best example I know of where a computer played a major role in proving a mathematical theorem is the Four Colour Theorem.

In 1852, shortly after he completed his studies at University College, London, Francis Guthrie wrote to his brother Frederick, still a student at the college, pointing out that as far as he could see, every map drawn on a sheet of paper can be coloured in using only four colours, in such a way that any two countries which share a stretch of common border are coloured differently (a feature which is obviously desirable in order to distinguish the various countries). Francis wondered if there was some mathematical proof of this fact – if fact it was. Frederick passed on the problem to his professor, the famous mathematician Augustus De Morgan.

Although not able to solve the problem, De Morgan did manage to make some progress on it. For instance, he proved that in any map it is not possible for five countries to be in a position such that each of them is adjacent to the other four. At first glance this would appear to solve Guthrie's problem, but a few moments' thought ought (!) to indicate that it does not. (Though over the 124 year period between the posing of the problem and its final solution, a period in which the Four Colour Problem, as it became known, grew in notoriety, numerous amateur mathematicians, upon rediscovering De Morgan's result, thought that they had thereby solved the problem.)

In common with practically anyone who has worked on the Four Colour Problem, we should begin by noting two basic facts. Firstly, there are simple maps which cannot be coloured using only three colours. Figure 2 gives an example of such a map.

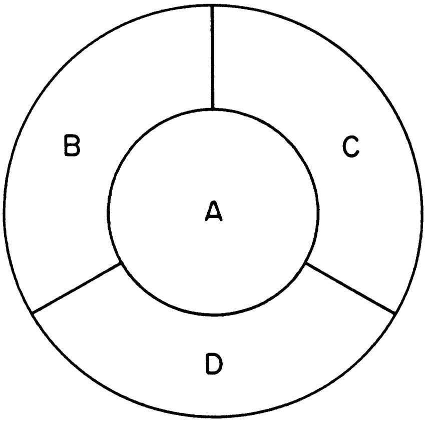


*Figure 2.* A simple map which requires four colours in order to be coloured so that countries which share a common stretch of border are coloured differently

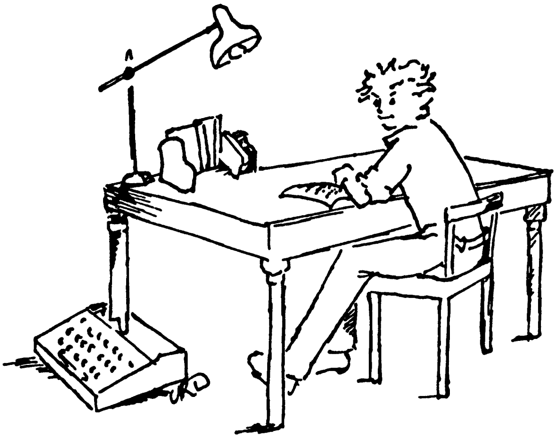
Secondly, five colours suffice for any map. This second result is a simple consequence of De Morgan's theorem, mentioned a moment ago, about which it should be said that, though it does not solve the problem, it is nevertheless a powerful result. It is powerful because it deals with *any* map, not just some particular maps, however complicated they may be. This is one of the great difficulties about the Four Colour Problem: it asks about all possible maps, of which there are infinitely many. Even a computer cannot handle infinitely many objects. (Actually, that use of the word 'even' is a bit silly, but we get so used to hearing about the power of computers these days that it is easy to slip into thinking about them as somehow 'all-powerful', which they most certainly are not, of course.)

Well, my last remark notwithstanding, in 1976 the Four Colour Problem was solved (thereby becoming the Four Colour Theorem), and the proof did involve the essential use of a computer (three computers, in fact). The credit for the proof has to be spread over three contributors: the mathematicians Kenneth Appel and Wolfgang Haken, and their computer(s). Neither party, the mathematicians nor the computer, could have completed the proof alone; each played a crucial role in the game. All of the mathematical ideas involved in the proof were supplied by mathematicians, but the proof involved such lengthy calculations that no human could ever follow them all, and these had to be left to the computer.

The central idea behind the proof goes back to a London barrister and amateur mathematician called Alfred Bray Kempe, who, in 1879, produced what turned out to be a false proof of the Four Colour Theorem, but one whose central strategy is essentially correct. What Kempe did was this. He reduced the problem to two separate problems. First of all he showed that any map which requires five colours has to contain one or more of a certain collection of special configurations of countries. Then, quite separately, he showed that none of the special configurations could in fact occur in a map which required five colours. Taken together, these two results clearly imply that four colours will suffice for any map. Unfortunately, Kempe's proof contained a sizeable hole: his collection of special configurations was not large enough to allow for all possible maps. This turned out to be not surprising, for Appel and Haken discovered that you must look at some 1500 different arrangements to make the proof work!

In 1976, then, after some 1200 hours of computer time, it finally proved possible to carry through Kempe's original strategy, using the computer to list and examine each of the 1500 special map configurations necessary for a correct analysis. It should be stressed that it was not simply a matter of programming the computer merely to run through all cases. Rather the computer and mathematician worked together, computer output leading to a response from the mathematician, and that in turn leading to more computation. So the result is a genuine product of the combined effort of man and machine.

Since the final proof of the Four Colour Theorem was not something which a mathematician could simply sit down and read – it was far too 'long' for that – many mathematicians at the time refused to acknowledge it as a 'proof' at all. By and large this view no longer prevails, and it is agreed that it is enough to read the computer program which carries out the calculations. The computer is now accepted as a legitimate tool *within* a mathematical proof. Which means that for the first time in the history of mathematics, the nature of what constitutes a mathematical proof has been modified. Whether this modification is a large one or not depends upon your point of view.



THE MICROCOMPUTER CAN BE OF GREAT ASSISTANCE TO TODAY'S MATHEMATICIAN

### *The Biggest Computer in the World*

It weighed 300 tons, and took up a total of 20,000 square feet of floor space, with an equal amount of space taken up by various peripherals. It was delivered to the United States Air Force in the late 1950s. The 18 removal vans it came in took 3 days to unload, to say nothing of the 35 vans containing the peripherals and spare parts. 'It' was the IBM AN/FSQ-7, the largest computer the world has ever seen, which was designed to run the US Air Force air defense system, SAGE ('Semi-Automatic Ground Environment'). The Air Force in fact bought 56 of these $30 million monoliths.

The entire system was designed by the Massachusetts Institute of Technology (at a special institute formed for the purpose), and an entire corporation was founded to write the software. The total bill for the network was around $8 billion. The program occupied 3 million punched cards. The hardware included 170,000 diodes and 56,000 vacuum tubes. Each installation in the network contained enough electrical wiring to stretch across the entire United States.

It was the first system to use interactive graphics displays and the first to employ data transmission to and from remote sites. It was not fully decommissioned until 1983, which also makes it the world's longest lived computer, a record it is likely to retain for all time judging by today's turnover!

# 6 The Oxen of the Sun (or how Archimedes' number came up 2000 years too late)

*"Compute, O friend, the number of the oxen of the Sun,*
*giving thy mind thereto, if thou hast a share of wisdom."*

Thus begins an epigram written in the third century B.C. by the famous Greek mathematician Archimedes, and communicated to Eratosthenes and his colleagues in Alexandria. The epigram goes on to describe an arithmetical problem involving the determination of the number of cattle in a certain herd, starting from nine stated constraints. The epigram also states that one who can solve the problem would be "not unknowing nor unskilled in numbers, but still not yet to be numbered among the wise." Nothing could be more apt, as it turns out. The problem was not solved until 1965, when a computer was brought to bear on the problem. The solution is a number having 206,545 digits! Clearly, Archimedes cannot himself have known the solution, but the wording of the epigram makes it clear that he knew it had to be pretty big. Doubtless he had quite a chuckle at the thought of the poor Alexandrians trying to find the solution.

'The Cattle Problem', as it is sometimes known as, has to do with a herd of cattle, consisting of both cows and bulls, each of which may be white, black, yellow or dappled. The numbers of each category of cattle are connected by various simple conditions. To give these, let us denote by $W$ the number of white bulls, and by $w$ the number of white cows. Similarly, let $B$, $b$ denote the number of black bulls and black cows, respectively, with $Y$, $y$ and $D$, $d$ playing analogous roles for the other colours. Using Archimedes' method of writing fractions (that is, utilising only simple reciprocals), the first seven conditions which these various numbers have to satisfy are

(1) $W = (\frac{1}{2} + \frac{1}{3})B + Y$

(2) $B = (\frac{1}{4} + \frac{1}{5})D + Y$

(3) $D = (\frac{1}{6} + \frac{1}{7})W + Y$

(4) $w = (\frac{1}{3} + \frac{1}{4})(B + b)$

(5) $b = (\frac{1}{4} + \frac{1}{5})(D + d)$

(6) $d = (\frac{1}{5} + \frac{1}{6})(Y + y)$

(7) $y = (\frac{1}{6} + \frac{1}{7})(W + w)$

The two remaining conditions are

(8) $W + B$ is a perfect square (that is, equal to the square of some number)

(9) $Y + D$ is a triangular number (that is, equal to a number of balls, say, which can be arranged in the form of a triangle, which is the same as saying that the number must be of the form $\frac{1}{2}n(n + 1)$ for some number $n$).

The problem is to determine the value of each of the eight unknowns, and thence the size of the herd. More precisely, what is sought is the *least* solution, since the conditions of the problem do not imply a unique solution.

If conditions (8) and (9) are dropped, the problem is relatively easy, and the answer was presumably known to Archimedes himself. The smallest herd that will satisfy conditions (1) to (7) consists of a mere 50,389,082 oxen. But the presence of the additional two conditions make the problem considerably harder. In 1880, a German mathematician called A. Amthor showed that the total number of cattle was a 206,545 digit number beginning with 7766. (If you want to find out how he was able to figure this out, you will have to look at his original writing on the subject, to be found in the scientific journal *Zeitschrift für Mathematik und Physik (Hist. litt. Abteilung)* 25 (1880), page 156.) Over the following 85 years, a further 40 digits were worked out.

It has been claimed that the first complete solution to the problem was worked out by the Hillsboro (Illinois) Mathematical Club between 1889 and 1893, though no copy of their solution exists as far as I know, and there is some evidence that what they did was simply to work out some of the digits and provide the algorithm for continuing with the calculation. At any rate, in 1965, H. C. Williams, R. A. German and C. R. Zarnke at the University of Waterloo in Canada used an IBM 7040 computer to crack the problem, a job which required $7\frac{1}{2}$ hours of computer time and 42 sheets of print-out for the solution. In 1981, Harry Nelson repeated the calculation using a CRAY-1 computer. This record-breaking machine required only 10 minutes to produce the answer, which was published in *Journal of Recreational Mathematics*, 13 (1981), pages 162–176. (The computer print-out is photoreduced to fit into 12 pages of the article.) The existence of this published copy of the answer at least saves me the task of giving it here.

What I will do is finish this chapter with another quotation from Archimedes. Archimedes was the son of Pheidias, a leading astronomer of the time, and was a good friend of (if not related to) king Hieron and his son and eventual successor Gelon. One of Archimedes' works, 'The Sand Reckoner', takes the form of a letter to king Gelon. In it, Archimedes calculates the number of grains of sand it would take to fill the entire visible universe. The answer he finally arrives at is $10^{63}$ (give or take the odd grain). Along the way, alongside some standard looking geometrical arguments, he observes that 10,000 grains of sand would occupy a volume equivalent to that of a poppy seed, and that the diameter of a poppy seed is at least one-fortieth of a finger-breadth, facts which, no matter how accurate, would be hardly likely to find their way past the editor of a modern mathematical journal (more's the pity). After arriving at his answer, Archimedes finishes thus

> *"I conceive that these things, king Gelon, will appear incredible to the great majority of people who have not studied mathematics, but that to those who are conversant therewith and have given thought to the question of the distances and sizes of the earth, the sun and moon and the whole universe, the proof will carry conviction. And it was for this reason that I thought the subject would be not inappropriate for your consideration."*

### *Computer Users Beware*

The story of the Oxen of the Sun told in the previous chapter should serve as a warning that it is not always a good idea to attack a mathematical problem by writing a computer program to do all the work. Even simple-looking problems can sometimes lead to quite enormous numbers. It is usually prudent to start off by taking a careful look at the problem. Even though a computer was required in order to find out how many oxen were in Archimedes' herd, this came at the end of a considerable mathematical analysis of the problem. A naive trial and error searching procedure would never stand a chance of coming up with a number of over 200,000 digits!

Another splendid example of simple-looking problems leading to incredibly large answers is provided by a cute little result due to the famous English mathematician D. E. Littlewood. Littlewood asked for the smallest number, $N$, with the property that if you shift the first digit of $N$ to the end, the result is exactly the original number $N$ multiplied by 3/2.

There is an obvious way of programming a computer to do this, but I do not recommend it. (Though you could ask an unsuspecting friend to try it!) The answer is a number in excess of a thousand million million. Using a clever trick of Littlewood, it is an easy matter to calculate this number.

Let the number $N$ be $a_1a_2 \ldots a_r$; that is, $a_1, a_2$, up to $a_r$ are the digits of $N$. Look at the fractional number

$$x = a_1.a_2a_3 \ldots a_ra_1\ a_2a_3 \ldots a_ra_1 \ldots\ldots$$

(The first dot here is a decimal point. The others indicate missing digits. So, in particular, the decimal part of the number $x$ goes on for ever in the repetitive pattern indicated.)

If we subtract $a_1$ from $x$ we get the number

$$x - a_1 = 0.a_2a_3 \ldots a_ra_1\ a_2a_3 \ldots a_ra_1 \ldots\ldots$$

The property of $N$ that we are looking at says that

$$a_2a_3 \ldots a_ra_1 = \tfrac{3}{2}(a_1a_2 \ldots a_r)$$

So

$$x - a_1 = \frac{3}{2} \cdot \frac{x}{10}$$

Rearranging this formula gives the equation

$$17x = 20a_1$$

So $x = \frac{20}{17}a_1$. For the smallest possible value of $x$, and hence of $N$, we should take $a_1 = 1$. This gives

$$x = \tfrac{20}{17} = 1.1764705882352941\ 1764705882352941 \ldots\ldots$$

So the number we are after is

$$N = 1{,}176{,}470{,}588{,}235{,}294$$

# 7 100 year old problem solved

In the autumn of 1983, a famous problem of mathematics that had resisted all attempts at solution for over 100 years was finally solved. Such is the nature of mathematics that, though famous among mathematicians, this problem was hardly known to the general public, though in many respects it is more fundamental than the widely known Fermat's Last Theorem problem, to which it is somewhat related.

Known as the 'Class Number Problem', the problem concerns the most basic of all mathematical notions, that of number. The final solution, by Benedict Gross and Don Zagier in the United States, required considerable knowledge of vast areas of modern pure mathematics, most of which is not known to the majority of trained mathematicians, let alone to the interested outsider. Fortunately, however, it is quite easy to say what the problem itself involves.

In our everyday lives we make use of just two kinds of number. There are whole numbers (positive and negative), known to mathematicians as 'integers', and fractional numbers, known in the trade as 'rationals'. For each type of number, if you add, subtract, or multiply two numbers of that type the result is another number of the same type. (Division has to be excluded here, since the result of dividing one whole number by another is not always a whole number.) Scientists and engineers frequently have to deal with two further types of number. First there are the 'real numbers'. These include the integers and rationals, as well as other numbers, like $\pi$, which, although not rational, can be approximated to any desired degree of accuracy by rational numbers. Then there is a still richer class, the 'complex numbers'. These are numbers of the form $a + bi$, where $a, b$ are real numbers and $i$ is the square root of minus-one.

To the outsider, the whole idea of there being a square root of minus-one seems absurd, and this reaction is not helped by the fact that for historical reasons the square root of minus-one is said to be an 'imaginary number', though in point of fact it is neither more nor less an imaginary quantity than is $\pi$. But strange or not, complex numbers are very useful. For instance, they are used in the mathematical theory of the alternating electric current that is used to light our homes. As with the integers and the rationals, if you add, subtract or multiply two real or complex numbers, the result is another number of the same kind.

So now we know of four quite distinct types of 'number' that are in common use. This is by no means the end of the story. In fact it is only the beginning. Mathematicians have, over the years, studied many, many different kinds of 'number', some relatively obscure, but some of considerable use within the field of mathematics. Indeed, for the mathematician, the study of new kinds of number is irresistible: for who else has the opportunity to recreate in maturity the wonderment of the child when first coming to understand the nature of numbers and their properties?

In the nineteenth century, the famous German mathematician Karl Friedrich Gauss studied a system of numbers now known as the 'Gaussian integers'. These are numbers of the form $m + ni$, where $i$ is the square root of minus-one (as in the complex numbers) and $m$ and $n$ are integers. Gauss was able to demonstrate that these numbers have many of the properties that we associate with the (ordinary) integers. For instance, if you add, subtract or multiply two Gaussian integers the result is another Gaussian integer. Also, in the Gaussian integer arithmetic it is possible to define the notion of a 'prime number', and to show that every Gaussian integer can be divided up into a product of 'primes', just as is the case with the ordinary integers. This means that the Gaussian integers have an arithmetic which in many ways is very similar to the everyday arithmetic of whole numbers: though as you might imagine, with the square root of minus-one floating around there are some surprises in store as well.

The Gaussian integers turned out to be so useful that Gauss himself began to look at other, similar kinds of number. In particular he considered numbers of the form $m + nk$, where $k$ is the square root of some fixed negative number other than minus-one. Now things get really interesting. Gauss discovered that these number systems have the above-mentioned properties of numbers, including the notion of primes and the factorisation of numbers into a product of primes (in a unique way), if $k$ is taken to be the square root of any one of

$$-1, -2, -3, -7, -11, -19, -43, -67, -163$$

The crucial point here is the uniqueness of any factorisation into primes. For instance, if $k$ is the square root of $-5$ (not in the above list) then the number 6 (that is, $6 + 0\,k$) has the two factorisations $2 \times 3$ and $(1 + k) \times (1 - k)$. In 1934, Heilbronn and Linfoot demonstrated that there could be at most ten values of $k$ which would give a number system which had unique factorisation into primes. Since there are nine in the above list, this led to a search for the missing number. By 1967 it was known that there was no tenth number: the above list is complete.

Now the attention was turned to another question that Gauss had started to look at. For each value of $k$, it is possible to associate with the corresponding number system a positive whole number called the 'class number' of that system. The actual definition of this quantity is somewhat abstract (even by the present standards!), but what this number does is give a measure of just how badly unique prime factorisation fails, if it fails at all. For the numbers in the above list, the class number is 1, which says that factorisation *is* unique. If the class number is 2, for instance when $k$ is the square root of $-5$ or of $-6$, unique factorisation 'only just' fails. The larger the class number is, the greater the degree by which unique factorisation fails and the less like the ordinary integers the appropriate number system looks.

The list quoted above gives all values of $k$ which give a number system with class number 1. Gauss himself looked at the problem of listing all values of $k$ which give a number system with class number 2, with class number 3, and so on. But neither he nor anyone else made much progress until 1967, when Harold Stark of MIT and Alan Baker of Cambridge found all values of $k$ which give a class number of 2. Then, in one fell swoop, Gross and Zagier solved the entire problem, showing just which values of $k$ give any prescribed class number. Thus, just 100 years after it all began, mathematicians now have the complete picture.

## *How to Find the Square Root of Computer, and Other Problems*

Take the word COMPUTER. Assign a digit to each letter in this word, each letter receiving a different digit. This gives you an eight digit number. It may or not be a perfect square. If it is, take its square root. This will be a four digit number. Its digits may or may not be distinct digits which were all assigned to the letters of COMPUTER. If they are, when you replace the digits of your four digit number by their corresponding letters you may or may not get a word of the English language. In fact there is only one way of assigning the digits to COMPUTER so that you do end up with a four letter word. (Curiously enough, it is a word used all the time by computer users.) What is this assignment of digits and what is the final four letter word? Your micro should take some time to sort that out, even with your help at the very end.

******

Arrange the digits 1 to 9 into two numbers so that one is the square of the other. All the digits must be used and no digit may be used twice. There are two solutions, and it should not take your micro long to find them both, especially if you spend a few moments thinking about the problem before you write your program. (If you have the patience, you should be able to find the two solutions without any aid at all. It takes about an hour I am told!)

******

Find three different perfect squares such that one of them is the average of the other two.

******

Which ten digit perfect square contains the most zeros? How about the most ones? Or the most twos? And so on.

******

Use the digits 0 to 9 to construct two numbers, the first one the cube of a number $N$, the second the fourth power of $N$. There is only one way to do this so that each digit is used exactly once. What is $N$ and what are the two numbers? A few minutes' thought before you write your program and this one should be easily disposed of.

# 8 Mod mathematics 1801 style

Modular arithmetic is something we become familiar with at a very early age. We use it as soon as we learn to tell the time. On a standard clock, we count hours *modulo* 12, which means that as soon as we reach 12 we start again at 1. (A more sophisticated way of describing this is to say that in mod 12 arithmetic the numbers 12 and 0 coincide.) On a 24 hour clock, hours are counted modulo 24. On either clock, minutes are counted modulo 60, as are seconds. Days of the week are counted modulo 7, months modulo 12.

The mark of a true genius is to see that something which everyone takes for granted embodies a profound and far-reaching concept. This is what the famous German mathematician Karl Friedrich Gauss did with modular arithmetic in 1801. Then only 24 years old, Gauss published his first book, *Disquisitiones Arithmeticae*, in that year; in this book he developed the idea of modular arithmetic into a powerful tool which has tremendous ramifications in modern day computational mathematics.

Gauss defined two numbers $m$ and $n$ to be *congruent modulo* a number $d$ if the difference $m - n$ is exactly divisible by $d$. An alternative, equivalent way of saying the same thing is that $m$ and $n$ are congruent modulo $d$ if both $m$ and $n$ leave the same remainder on division by $d$. Gauss used the notation

$$m \equiv n \pmod{d}$$

for this concept, a notation still used today.

For instance, 27 is congruent to 12 modulo 5, since both leave a remainder of 2 upon division by 5. (Alternatively, $27 - 12 = 15$, which is divisible by 5.) The remainder upon dividing a number $n$ by $d$ is called the *residue* of $n$ modulo $d$, and it is obvious that every number is congruent modulo $d$ to its residue modulo $d$. In modular arithmetic for some chosen modulus $d$, only residues are considered. In mod 12 arithmetic, the residues are the numbers 0 to 11. When you reach 12, you immediately replace it by 0 and start again.

Except for the convention that we use the number 12 rather than 0, this is exactly what we do when we tell the time.

There are simple rules for adding and multiplying residues in mod $d$ arithmetic. You add and multiply in the usual way, but discard any multiple of the modulus $d$ that arises. For instance, to add the residues 5 and 6 modulo 8, you first add 5 and 6 to get 11 in the usual way, and then throw away 8 from the 11 to leave 3, the answer. To multiply 5 and 6 modulo 8 you first multiply 5 and 6 in the usual way to get 30, and then throw away 24, the largest multiple of 8 less than 30, to leave the answer 6, which, curiously enough is one of the two numbers that you started with. (This is just one of the surprises that are in store for anyone who plays around with modular arithmetic.)

Subtraction of residues is also possible, in much the same way that addition and multiplication are, but with a new twist. In ordinary subtraction you very often get a negative answer, as when you perform the subtraction $5 - 8 = -3$. In modular arithmetic there are no negative numbers. For instance, if you want to subtract 8 from 5 modulo 11, then $5 - 8 = -3$, but the residue of $-3$ modulo 11 is 8 (because $8 - (-3) = 8 + 3 = 11$), so this is the answer. Division is sometimes possible, as with ordinary arithmetic of whole numbers. In ordinary arithmetic, you can divide 6 by 3 but cannot divide 6 by 4 (and get a whole number for an answer.) In modular arithmetic the same phenomenon occurs, but again there is a twist. If the modulus you are using is a prime number (that is, if it has no divisors besides itself and 1) then division is always possible. For instance, suppose you want to divide 14 by 6 modulo the prime 17. In ordinary arithmetic this is not possible, but in mod 17 arithmetic the answer is 8. How do I know? Well, how do you check the answer from any division sum? You multiply back up. If the result of dividing 14 by 6 really is 8, when you multiply 6 by 8 you should get the number 14. Well, $6 \times 8 = 48$, and modulo 17 the number 48 is the same as 14 since two lots of 17 (giving 34) must be discarded.

Arithmetic of residues obeys the same rules as does ordinary arithmetic. For instance, the order in which you add or multiply any number of residues is unimportant, and multiplying $M$ by $N$ gives the same answer as adding $M$ to itself $N$ times. Because of this, Gauss was able to demonstrate some rather nice little arithmetic rules for seeing whether a certain number is exactly divisible by another in regular arithmetic. The connection between regular division and residue arithmetic is the observation that saying the number $N$ is divisible by the number $M$ is the same as saying $N$ is congruent to 0 modulo $M$. Using this observation, it is not at all difficult to show that a number $N$ will be divisible by 3 if, and only if, the sum of the digits is divisible by 3. For instance, is the number 139842186321 divisible by 3? Well, if you add together all the digits of this number you get 48, which is divisible by 3, so the original number is divisible by 3 as well. For larger numbers, this trick can obviously save having to perform a hideous long division sum. Another similar result is that a number is divisible by 9 if, and only if, the sum of its digits is divisible by 9. Similar tricks are used in present day computer programs designed to factorise large numbers, a problem of great interest because an efficient method of factorising large numbers would provide cryptographers with a method for cracking the so-called *Public Key Cryptography* codes (see chapter 13) in widespread use by the world's security organisations.

Gauss' congruence arithmetic also provides mathematicians with a slick method for proving the famous 'Little Theorem' of the seventeenth century French genius Pierre De Fermat. This says that if $p$ is any prime number, then for any number $N$, the number $N^{p-1} - 1$ is always divisible by $p$. In Gauss' notation this says that

$$N^{p-1} \equiv 1 \pmod{p}$$

This result, besides being the reason why many predictive card tricks work, also forms the basis of all modern primality testing methods.

Since a prime is defined to be any number which is not divisible (without remainder) by any number other than itself and 1, the obvious method for testing to see if a given number is prime is to test if any smaller numbers divide it. (In fact, it is enough to see if there is any prime less than the square root of the number which divides it.) Though this method works well for reasonably small numbers, it is not at all feasible for large ones. For instance, to check if a number of just 50 digits were prime using this method could take a hundred billion years, using the fastest computers available. Using an alternative method based upon Fermat's Little Theorem the same kind of computer would take just 15 seconds at the outside. And this is not an irrelevant observation. The *Public Key Cryptosystems* mentioned earlier use prime numbers with 100 digits.

Suppose you want to check if a number $N$ is prime using the Fermat Theorem. This tells you that if $N$ is prime, then $a^{N-1} \equiv 1 \pmod{N}$ for all numbers $a$ less than $N$. So if you can find a number $a$ for which $a^{N-1}$ is not congruent to 1 modulo $N$, you will know at once that $N$ is not prime. But how do you look for such a number $a$? In theory, you could simply look through all possibilities. If at the end you have tested all numbers $a$ less than $N$ and found that $a^{N-1} \equiv 1 \pmod{N}$ for all of them, then you can be sure that $N$ is prime. If you find an $a$ for which $a^{N-1}$ is not congruent to 1 modulo $N$, you will know that $a$ is not prime. So, in theory this method will work. The trouble is that it is even more inefficient than trial division.

Efficient primality tests based upon the Fermat Theorem work by looking at a *very small number* of test values for the number $a$. For instance, if you take the number 2 and work out the residue modulo $N$ of $2^{N-1}$, and the answer is 1, then it is highly likely that $N$ is prime. For instance, against the 5,761,455 primes less than one hundred million, only 2057 non-primes will give $2^{N-1} \equiv 1 \pmod{N}$, so if you simply use this calculation as a primality test the probability of error is only about one in three-thousand. But as you may imagine, there are many instances where such a probability of error is unacceptable. In such cases, use is made of sophisticated techniques for eliminating all possibilities of error from this procedure. The mathematics required to do this is by no means simple, and the test you end up with looks quite complicated (though still extremely efficient from a computational point of view). Nevertheless, at heart all that it is is a souped up version of the Fermat Little Theorem.

## *How to Find Record Primes*

The largest known prime number at the time of writing is the 39,751 digit giant

$$2^{132049} - 1$$

discovered using a CRAY-XMP computer in September 1983. Record primes are nowadays always of the form $2^n - 1$, since for numbers of this type there is an extremely efficient primality test called the Lucas–Lehmer Test.

To test if a number $2^n - 1$ is prime you perform the following calculation. First check to see if $n$ is prime. If it is not, then $2^n - 1$ cannot be prime. If $n$ is prime, calculate the number sequence $U(0), \ldots, U(n-2)$ using the rules

$$U(0) = 4$$

$$U(k+1) = (U(k)^2 - 2) \bmod (2^n - 1)$$

(So you first calculate $U(1)$ from $U(0)$, then $U(2)$ from $U(1)$, and so on.) If at the end you find that $U(n-2) = 0$, then $2^n - 1$ is prime. If $U(n-2)$ is not equal to 0, then $2^n - 1$ is not prime (see the example given below).

Using this test it took the CRAY-XMP computer just over an hour to show that the record prime given above is in fact prime, so the home micro user cannot hope to compete at world record level, but it should be possible to perform the calculations on the world record up to, say, at least the early 1970s. A list of these numbers is given on page 102.

As an illustration, I will show that the number $2^5 - 1$ is prime. (Since $2^5 - 1$ is easily calculated to be 31, this is a fairly inefficient way of doing it in this case, but at least the numbers involved are small enough to make the calculation an easy one to perform without aid.) I have to calculate the numbers $U(0)$, $U(1)$, $U(2)$, $U(3)$. I get

$$U(0) = 4$$

$$U(1) = (4^2 - 2) \bmod 31 = 14 \bmod 31 = 14$$

$$U(2) = (14^2 - 2) \bmod 31 = 194 \bmod 31 = 8$$

$$U(3) = (8^2 - 2) = 62 \bmod 31 = 0$$

Since $U(3)$ is 0, the number $2^5 - 1$ is prime.

# 9 Another slice of pi

The mathematical constant $\pi$ ('pi'), defined to be the ratio of the circumference of any circle to its diameter, is known to be an *irrational* number, which is to say that its decimal expansion continues indefinitely, without settling down to any recurring pattern. The infinite decimal expression for $\pi$ begins

$$3.14159\ 26535\ 89793\ 23846\ \ldots\ldots$$

Computers have been used to calculate the number as far as 16 million decimal places, which is a lot of slices, whichever way you cut it (see chapter 2).

One of the amazing things about $\pi$ is that it keeps cropping up in the most unexpected places. For instance, suppose I asked you to calculate the value of the infinite sum

$$1 - 1/3 + 1/5 - 1/7 + 1/9 - 1/11 + \ldots$$

where those dots indicate that the sum continues indefinitely in the manner indicated (in this case keep alternating plus and minus signs and increasing the denominators of the fractions by 2 each time, to run through all odd numbers). Of course, you cannot calculate this sum by adding together all the individual fractions involved (taking account of signs), as there are an infinite number of them. (In point of fact, even a large finite number of terms would cause you a problem, but in the idealised world that is mathematics it is usually agreed that, in principle, you could work out the answer, however many numbers are involved.) But mathematicians have ways of calculating what the answer to such a sum *would* be *if you were* able to perform the infinitely many additions involved. And for the sum quoted, the answer is $\pi/4$.

Another example is the infinite sum

$$1 + 1/4 + 1/9 + 1/16 + 1/25 + \ldots$$

Here the individual terms in the sum are the reciprocals of the squares of the numbers 1, 2, 3, 4, 5, etc. The answer in this case is $\pi^2/6$. There are many other infinite sums (and products) whose answers involve $\pi$ (see chapter 2). In fact, it is the existence of such sums which enables mathematicians to calculate $\pi$ to many decimal places. Certainly, drawing circles and measuring them would not get you far. A mere 10 decimal places would enable you to calculate the circumference of the Earth from its diameter to an accuracy of a few inches.

The number $\pi$ makes another surprise appearance when you work out your chances of winning at the following simple game. A board has lines ruled on it in parallel, at a distance of 1 inch apart. You throw a 1 inch matchstick on to the board, and you win if the matchstick ends up not touching any line. Would you play this game if the stakes were, say, £50 a time? Well, the probability that the matchstick will end up touching a line is precisely $2/\pi$, or a little less than 2 out of 3, so your chances of winning are just over 1 in 3. You would not be wise to get involved in such a loaded game. The exact formula involved here is this. If the lines are a distance $d$ apart and the stick has length $l$, the probability that the stick will land touching a line is $2l/d\pi$. For instance, if the lines are 2 inches apart and the stick is 1 inch long, the stick lands on a line with probability exactly $1/\pi$, which would give you a game well worth playing.

And how about this for an unexpected appearance of $\pi$. If you choose two whole numbers at random, the probability that they have no common factor (other than 1, which is a factor of all numbers) is $6/\pi^2$. It is an easy matter to check this result, but if you are one of those people who hates mathematical manipulations, just skip over the next paragraph.

Let $P$ denote the unknown probability that two numbers chosen at random have no common factor. Now, if you take any two numbers $A$ and $B$, the highest common factor of $A$ and $B$ will be equal to $N$ if, and only if, both $A$ and $B$ are multiples of $N$ and the numbers $A/N$, $B/N$ have no common factor (other than 1). The probability that a number chosen at random is a multiple of $N$ is obviously $1/N$. (Think about it a minute.) So the probability that $A$ and $B$ have highest common factor $N$ is

$$1/N \times 1/N \times P$$

Since any two numbers must have a highest common factor (if we include 1 as a possibility), if we add together all of the probabilities worked out above for all values of $N$ the answer has to be 1 (that is, certainty). So

$$(1/1 \times 1/1 \times P) + (1/2 \times 1/2 \times P) + (1/3 \times 1/3 \times P) + \ldots = 1$$

Taking out the common factor $P$, this becomes

$$P \times [(1/1 \times 1/1) + (1/2 \times 1/2) + (1/3 \times 1/3) + \ldots] = 1$$

In other words

$$P \times [1 + 1/4 + 1/9 + \ldots] = 1$$

But we know what the answer is for the infinite sum in the brackets here. As mentioned above, it is $\pi^2/6$. Hence $P = 6/\pi^2$, as claimed.

The fact that $\pi$ is irrational was first proved by Lambert in 1770. Somewhat deeper is Lindemann's result of 1882 that $\pi$ is in fact *transcendental*, which as far as mathematicians are concerned means that $\pi$ is not a solution of any polynomial equation with rational coefficients. (So, for example, there is no quadratic equation of the form

$$ax^2 + bx + c = 0$$

where $a$, $b$, $c$ are rationals, which has $\pi$ as one of its roots.) One consequence of this is that the Ancient Greek problem of *Squaring the Circle* using ruler and compasses alone is definitely impossible. (Not that the conclusive proof of this fact has deterred countless would be circle squarers from subsequently trying to 'solve' the problem. Erroneous solutions continue to be sent in to mathematicians around the world.)

The task posed by the Squaring the Circle problem is to find a method for constructing a square whose area is equal to that of a given circle. Now, if that were all that was involved, there would be no great problem. There are various methods of achieving this. But the problem specifies that the only two tools that are to be used are a ruler (more precisely, a straightedge, since no markings are allowed on it) and a pair of compasses (which have the property that they do not hold their angle when lifted off the paper). The use of these most basic of tools goes back to Plato's assertion that the only 'perfect' geometrical figures are the straight line and the circle.

As anyone who has done any Euclidean geometry at school (which means just about anyone now over the age of about 30) will be able to testify, using just these two tools, it is possible to construct a remarkable number of geometric figures. In fact, so versatile do these weapons prove to be, it is not surprising that the Ancient Greeks were astonished to find that they could not solve the circle-squaring problem using them. Of course, you may ask what all the fuss is about. If the problem can be solved quite easily using other tools, why bother to continue over the centuries with attempts to solve it in the manner required? A similar question would be why all the fuss about running a four minute mile when the distance can easily be covered in that time on a bicycle? Just as the Ancient Greeks liked their athletics, so too they loved intellectual challenges, and 'ruler and compasses' constructions, as they are known, provided a good game with simple, well-defined rules, in which anyone could participate.

The final impossibility proof for Squaring the Circle followed Lindemann's result that $\pi$ is transcendental. By introducing Cartesian coordinates into your figure and considering the coordinates of points which can be constructed as the intersections of circles and/or lines drawn in a ruler and compasses construction, it is a relatively easy task to show that if the circle could be squared, the number $\pi$ would be the solution to a polynomial equation with rational coefficients, which is precisely what Lindemann showed is not the case.

It is interesting to note that, simple though they are, a ruler and compasses constitute more than you really need for geometrical constructions. In 1672, Mohr showed that you can dispense with the ruler altogether, provided that you regard a line as drawn when you have specified two points on it. (In other words, the ruler is not necessary for the actual *construction*, only for making the figure intelligible to us.) And in 1883, Steiner showed that a ruler alone is enough provided you start off with one circle and its centre. In both of these cases, what we are saying is that these restricted tools will allow you to accomplish any construction possible using ruler and compasses, of course. Tasks such as Squaring the Circle remain as impossible as ever. It is also known that all ruler and compasses constructions are possible using just a *two-sided* ruler whose sides are parallel. (The fact that practically any ruler has two parallel sides is irrelevant here. In keeping with the Greek tradition, the 'ruler' used in 'ruler and compasses' constructions must be regarded as having only one usable edge.)

Just in case none of this convinces you, and you are determined to press on and square the circle, can I make one plea? Please do not send me your solution. I won't read it! Absolutely not! And that's definite!

### *Squaring the Circle*

If you have read the previous chapter, you will know that the Ancient Greek problem of Squaring the Circle cannot be solved in the manner stipulated. But if you still want to try your hand at circle-squaring, why not try the following. Can you square the circle in a linguistic sense? In other words, can you fill in the table below so that every row and every column is a genuine English word?

C I R C L E
I • • • • •
R • • • • •
C • • • • •
L • • • • •
E • • • • •

******

### *Squares and Cubes*

In keeping with the previous problem, try 'squaring the square'. And when you have done that, have a go at 'cubing the cube' (that is, starting with the word CUBE written in three perpendicular directions from the same initial letter C, construct a 4 × 4 × 4 letter cube which spells out a word in all possible directions.) And for real insomniacs, there is always hypercubing the hypercube, of course, a problem that is really out of this world.

# 10 Coincidence?

What are the three most common special 'constants' of mathematics? The most obvious one is $\pi$, of course, the ratio of the circumference of any circle to its diameter. The number $\pi$ is *irrational* (that is, its decimal expansion continues indefinitely without settling down to any repetitive pattern), beginning with

3.14159 26535 98793

The next most common one is the number $e$, the 'base of natural logarithms'. This number is also irrational, its decimal expansion beginning with

2.71828 18284 59045

There are several alternative definitions of the number e (as indeed there are for $\pi$). Possibly the most widely known is that e is the number you get when you calculate the answer to the infinite sum

$$1 + 1/1! + 1/2! + 1/3! + 1/4! + \ldots$$

The next most common one is the number e, the 'base of natural For any number $N$, the symbol $N!$ (read as '*N-factorial*') denotes the result of multiplying together all the numbers 1 to $N$, inclusively. So, for example, $2! = 1 \times 2 = 2$, $3! = 1 \times 2 \times 3 = 6$, $4! = 1 \times 2 \times 3 \times 4 = 24$. The only other point that perhaps needs some explanation is the use of those dots at the end of the sum. They indicate that the sum continues indefinitely in the manner indicated. So, for example, the next number to be added is 1/5! (that is, 1/120). Since the sum in question is an infinite one, it is impossible to work out the answer by adding in the usual sense, of course, so when mathematicians need to work with the exact value of e, they must use alternative means. But the sum can be used to work out the value of e to whatever degree of accuracy is required. The point is that the values of the factorial function grow large so rapidly (10! is over 3 million) that the successive terms in the sum quickly become less significant.

The letter e used for this number is from the word ‘exponential’. The function EXP($x$) available in most computer programming languages provides the values for the function ‘e raised to the power $x$’. We all know what e raised to the power 2, or 3, or 4, etc. means, but what about e raised to the power $\pi$, say? How can such a number be calculated? Mathematicians can answer that one immediately. For any number $x$, whole, fractional, or even irrational

$$e^x = 1 + x/1! + x^2/2! + x^3/3! + x^4/4! + \dots$$

One of the significant aspects of the function EXP($x$) is that, alone among all functions, if you draw its graph then the gradient of the slope at any point is exactly equal to the value of the function at that point: that is, the function tells you how fast it is growing at each point.

Just as $\pi$ keeps cropping up all the time in mathematics, so too does e. Pick up any textbook on statistics, for example, and you will find that it is littered with references to e. For instance, the equation which produces the familiar ‘normal distribution’ curve involves e.

The third common constant of mathematics is the ‘imaginary’ number, $i$, the square root of $-1$. Relatively few people take their mathematical education far enough to come across $i$, but in advanced parts of the subject it is used all the time. For instance, the mathematics of electric currents involves $i$, as any engineer will be able to tell you. To the uninitiated, this seems somewhat surprising. How can an ‘imaginary’ number be related to anything so concrete as electricity? The point is, the use of the word ‘imaginary’ for numbers such as $i$ is purely a historical accident. In reality, the number $i$ is no more ‘imaginary’ than are the numbers $\pi$ and e, each of which requires *infinitely many places of decimals* in order to be specified. And we all know that numbers like $\pi$ and e are useful. (At least if we have got this far we do.)

Now, on the face of it there is no reason why there should be any connection between the three constants $\pi$, e and $i$, is there? One has to do with circles, one with gradients of curves, and one with square roots of negative numbers. And yet, as Euler discovered, there is an amazing connection between these three numbers, namely the remarkable equation

$$e^{i\pi} = -1$$

Faced with this equation, you may ask just what is meant by raising a number to a power involving the number $i$? In the case of the number e, the answer is easily given. The formula given above for calculating $e^x$ can be used even if $x$ involves $i$. The *meaning* of imaginary exponentiation for numbers other than e is, I think best left alone just here, as with the *meaning* of irrational exponents such as $\pi$. Though I cannot resist telling you that it turns out that if you raise $i$ to the power $i$ the answer is real. In fact, it brings in $\pi$ and e again

$$i^i = e^{-\pi/2}$$

There are other, rather less dramatic connections between $\pi$, e, and $i$. For instance, if you replace $\pi$ by its equivalent Latin letter p, the letters p, $i$, and e together spell out the word 'pie'. No more significant is the fact that, to an accuracy of four decimal places

$$e^6 = \pi^5 + \pi^4$$

(The difference between $e^6$ and $\pi^5 + \pi^4$ is, to six decimal places

0.000017)

Having looked at the constants $\pi$, e, and $i$, what might the next most common constant be? There are several contenders. One which you will not find given as such in any mathematics book which I know of (except this one) is the number $\sqrt{163}$. As far as I know, there is no explanation for the fact that this particular number seems to crop up in strange places. It *may* all be pure coincidence. On the other hand . . . . Just see what I mean.

If you study the numbers of the form $m + ni$, where $m$, $n$ are whole numbers and $i = \sqrt{-1}$, they turn out to have many of the properties of the ordinary whole numbers. In particular there is a well-defined notion of 'prime number' of this sort, and every number like this can be factored in a unique way into a product of these primes. $\sqrt{-1}$ is not the only number that can be used here. There are more (see chapter 7) such numbers. The largest one is, wait for it, $\sqrt{-163}$, that is, $i\sqrt{163}$.

Euler discovered that the quadratic formula

$$f(n) = n^2 + n + 41$$

yields prime numbers as its value for all values of $n$ from 0 to 39 (see chapter 3). No other polynomial formula is known which comes remotely close to this. Moreover, out of the first ten million values of this function just under one-half are prime, a fact also unparalleled by other formulas known. Suppose now that you try to solve the quadratic equation

$$f(x) = 0$$

that is

$$x^2 + x + 41 = 0$$

Using the well-known formula for this you find that the solutions are

$$x = \frac{-1 \pm \sqrt{-163}}{2}$$

and there is that number $\sqrt{163}$ again.

Finally, if you calculate the number

$$e^{\pi\sqrt{163}}$$

you get the answer

262 53741 26407 68744 . 00000 00000 00

correct to 12 decimal places. Surely, when you raise the irrational number e to the irrational power $\pi\sqrt{163}$ the answer cannot be a whole number? Well, in fact it is not. A more accurate value is

262 53741 26407 68743 . 99999 99999 99250

So all that can really be said is that the number $e^{\pi\sqrt{163}}$ is very nearly a whole number. Now, in an exact science like mathematics, 'very nearly' is not a very meaningful concept. Either a number is a whole number or it is not. Period! But it is curious that the exponent which comes closest to achieving a whole number value when combined with $\pi$ and e in this manner is the mysterious $\sqrt{163}$. (Other values do not seem to come anything like as close as this one – at least not among the ones looked at so far.) It may of course all be coincidence. But then again it may not. With any luck, time will tell.

## *Irrational Numbers*

In the previous chapter it was mentioned that the numbers $\pi$ and e are *irrational*, that is, they require infinitely many decimal places to be specified exactly, the infinite decimal expression having no repeating pattern. Proofs that numbers are irrational can be quite difficult, involving lots of advanced mathematical techniques. For instance, although it is known that the number $e^{\pi}$ is irrational, this has not been proved for any of the numbers $e\pi$, $\pi^{\pi}$, $e^{e}$ or $\pi^{e}$.

******

If you multiply the number 1089 by 9 the answer is 9801, which is the original number written backwards. There is only one other number less than 10,000 which, when multiplied by another number, produces the original number back to front. What is it, and what is the multiplier?

And when you have solved this one, see if you can find any other examples beyond 10,000. If you can, let me know and thereby earn yourself immortality in any future editions of this book.

******

Find digits $A$, $B$, $C$ such that $ABC = A! + B! + C!$

# *11 Fermat's Last Theorem*

Early in 1983, the mathematical world was stunned by the news that a 29 year old German mathematician had obtained a partial solution to the famous problem of *Fermat's Last Theorem*. Exactly what Gerd Faltings of Wuppertal University did was to prove correct the *Mordell Conjecture*, a problem which had resisted all attempts at solution since its formulation in 1922. A precise statement of the Mordell Conjecture is not possible here, since it involves mathematical concepts not familiar to most people. What can be explained is how Faltings' result provides a partial solution to the *Fermat Problem* (as 'Fermat's Last Theorem' is known as to present day mathematicians).

Practically all of present day Number Theory can be traced back to the great French mathematician Pierre De Fermat (1601–65). To this day, Fermat serves as an inspiration to all amateur mathematicians. For he was a magistrate by profession, and had no formal mathematical training. In fact he was past 30 before he took up the subject at all, and then only as a pleasant pastime. Among his work can be found pioneering research which eventually led to the proper development of the differential and integral calculus, and some foundational work on probability theory, but it was in Number Theory that Fermat made his greatest advances.

Fermat himself published very little of his mathematical work – after all he only did it as a hobby. Most of what he did was carried out in the form of correspondence with other mathematicians, among whom were many of the leading professionals of the day. Often, when he had made a discovery he would simply announce the result, leaving its proof as a challenge to his contemporaries. Only very rarely did it turn out that he was in error (see chapter 15 for an instance of this rare phenomenon), or that a result he claimed was in fact not true, though on occasions over a hundred years were to elapse before anyone else was able to find a proof. (For instance, this was the case with the *Fermat Little Theorem* described in chapter 8.) Most of what we now know about Fermat's private researches stems from his habit of making short notes in copies of his textbooks, among which was a copy of the book *Arithmetica*. Written in Greek by the Alexandrian mathematician Diophantus around A.D. 250, only part of the original work still exists. It constitutes one of the first textbooks on Number Theory ever written. In honour of Diophantus, present day mathematicians use the phrase '*Diophantine equation*' to mean an equation whose coefficients are whole numbers, where a whole number solution for the unknowns is sought.

The surviving part of the *Arithmetica* became fully accessible to European scholars when Claude Bachet published a Latin translation in 1621. It seems likely that it was this work which first aroused Fermat's interest in Number Theory. Certainly his personal copy of the book became littered with his now famous marginal notes, and five years after his death his son Samuel published a new edition of the book incorporating his father's marginalia. By far the most notorious of these is the following (written in Latin)

> *To resolve a cube into two cubes, a fourth power into fourth powers, or in general any power higher than the second into two of the same kind, is impossible; of which fact I have found a remarkable proof. The margin is too small to contain it. . .*

What Fermat was saying is that for any number $n$ greater than 2, the equation

$$x^n + y^n = z^n$$

has no solution for $x, y, z$ among the whole numbers. (For $n = 2$, there are infinitely many solutions, the so-called *Pythagorean triples* such as 3, 4, 5 or 5, 12, 13.) The statement that such Diophantine equations have no solutions resisted attempts at a proof long after all of Fermat's other assertions were either proved or (rarely) disproved, so in time it became known as Fermat's *Last* Theorem, though it was by no means the last result he ever discovered. That Fermat had a correct proof of the assertion for the case $n = 4$ is not in doubt. But though every other assertion found in his *Arithmetica* turned out to be correct, there is very good reason to believe that on this occasion he made an error. Indeed, there is evidence that he subsequently discovered the error, but did not bother to amend his earlier margin note. (And why should he have bothered? The notes were never intended for any eyes but his own.) But in such a manner legends are born. The romantic idea of a seventeenth century amateur mathematician proving a result which resisted all attempts by professionals for over 300 years was bound to catch the public imagination. The more so when the problem is so easy to state. Consequently, while present day mathematicians tend to refer to Fermat's assertion as *The Fermat Conjecture*, it is still more widely known as *Fermat's Last Theorem*. (Incidentally, I do not think that the present day professionals are being particularly arrogant in this. There is an 'obvious' and very clever 'proof' of the Fermat Conjecture which fails because of an extremely subtle error which could not have been anticipated in Fermat's time, and was isolated only after several excellent mathematicians discovered practically the same false 'proof' in the middle of the last century.)

So just what was known about the Fermat Conjecture prior to Faltings' discovery? Well, in 1770, Euler proved the Conjecture for the case $n = 3$. (Actually there was a 'gap' in Euler's argument, which Legendre subsequently removed.) The case $n = 4$ had already been done by Fermat, as noted above. Some time around 1825, Dirichlet and Legendre independently settled the case $n = 5$, and in 1839 Lamé knocked off $n = 7$. (The case $n = 6$ follows directly from the case $n = 3$. In fact, it is easy to see that in order to prove the Fermat Conjecture all you need to do is prove it for the case where $n$ is a prime number. All other cases follow from this at once.) A real breakthrough was made by the German mathematician Kummer who, in 1844 proved that the Conjecture is true for all so-called *regular primes* (as the exponent). That this is a significant result can be gleaned from the fact that of the primes less than 100, only 37, 59, and 67 fail to be regular. Unfortunately, while computer research has demonstrated that 'almost all' primes are regular, it is still not known if there are infinitely many regular primes! (Computer evidence, though at times very helpful, falls far short of providing mathematical certainty – see chapter 15.) In fact, the situation is somewhat worse. In 1915, Jensen discovered that there is definitely an infinite number of non-regular primes. ("What value the computer evidence now?" you may cry!) More recently, mathematicians have shown that the Conjecture is true for a class of primes much more all embracing than the regular ones, and computers have been used to show that all primes less than 125,000 are in this larger class. So the Conjecture certainly is a Theorem for all exponents up to 125,000. (The precise definitions of both regular primes and the new larger class are outside the scope of this book.)

And so to Faltings. If you take the Fermat equation

$$x^n + y^n = z^n$$

and divide through by $z^n$ you get the equation

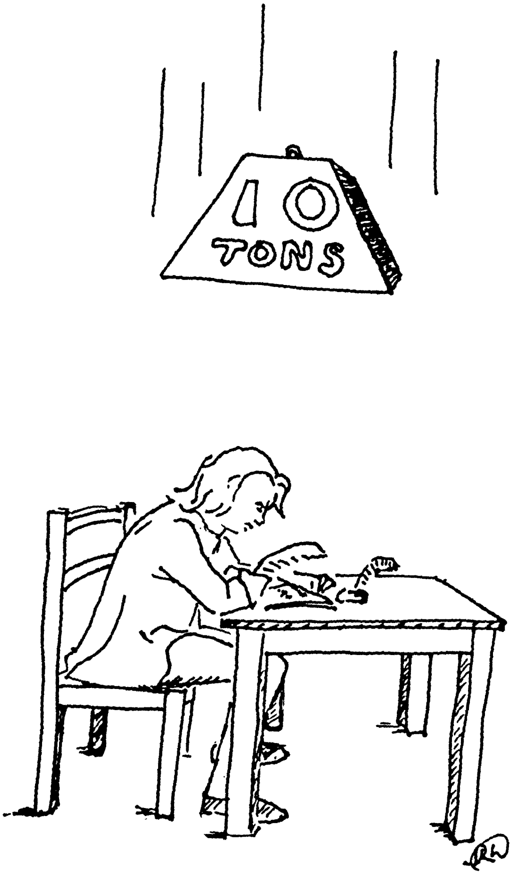
$$u^n + v^n = 1$$

where I have written $u$ instead of $x/z$ and $v$ for $y/z$. Since multiplying this new equation through by $z^n$ reproduces the original equation, this means that an equivalent way of stating the Fermat Conjecture is to say that for no $n$ greater than 2 does the equation

$$u^n + v^n = 1$$

have a solution for $u$, $v$ consisting of *rational numbers* (that is, fractions), except for the 'trivial' solutions where one of $u$, $v$ is 0 and the other is ±1. (It is a simple matter to demonstrate that any such equation will have an infinite number of non-trivial solutions which are not rational.) In 1922, Lewis Mordell conjectured that any such equation has only *finitely many* rational solutions. Certainly this is weaker than the Fermat Conjecture which says that there are *no* (non-trivial) rational solutions, but to a mathematician the step from the infinite to the finite is a very considerable one, so when Faltings managed to prove Mordell's Conjecture, he made a tremendous advance on the Fermat Problem. It may even lead to an eventual solution to the problem once and for all. Undoubtedly, any eventual solver of the Fermat Problem would thereby ensure his future, but the prize of 100,000 Marks for a solution offered by the Academy of Science at Göttingen in 1908 has long been rendered valueless by the German inflation of the 1920s.

Two final comments. As formulated by Mordell, his Conjecture applies to a much larger class of equations than the (modified) Fermat equations. It is the complexity in the description of this wider class of equations which prevents me stating the full conjecture here. Secondly, Faltings' proof is an indirect one. Even for the Fermat equations, it gives no idea as to how many solutions there may be, simply that their number is finite. So we do not really know *how* close we are to a final solution to the 300 year old teaser.



FERMAT'S LAST THEOREM

## *Hardy's Last Theorem?*

The origin of Fermat's Last Theorem described in the previous chapter brings to mind a story about the great English mathematician G. H. Hardy, who died in 1947. It involves a notorious unresolved conjecture called the Riemann Hypothesis. It would take too long to describe exactly what this says, but for the moment it is enough to know that either a proof or a disproof of this century old hypothesis would be one of the most significant results of present day mathematics, certainly much more significant than a proof of Fermat's Last Theorem.

Anyway, the story about Hardy goes like this. Hardy was due to return to England from Scandinavia by the North Sea Ferry, and was made more than a little apprehensive by the fierce weather that was prevailing at the time. Accordingly, just before the boat sailed he mailed a postcard to a colleague bearing the message 'Have proved the Riemann Hypothesis. Yours, G. H. Hardy.' He reasoned that God, knowing that Hardy was sufficiently reliable a mathematician for such a message to carry conviction, could not possibly let him perish with the undeserved credit of proving so important a result. At any rate, Hardy returned home safely.

# 12 Seven-up*

If you work out the reciprocal of the number 7 as a decimal you get the infinite repeating pattern

1/7 = 0.142857 142857 142857 . . . . . .

If you multiply this number by any of 2, 3, 4, 5, 6 (to obtain the decimal expressions for 2/7, 3/7, 4/7, 5/7, 6/7, respectively) a curious thing happens. Each time what you end up with is the same infinite repeating pattern shifted along a number of places. For example

2/7 = 0.2857 142857 142857 . . . . . .

and

3/7 = 0.42857 142857 142857 . . . . . .

Just how special is the number 7 in this respect? Are there any other numbers $N$ with the property that the decimal expression for $1/N$ is an infinite repeating pattern which is simply shifted along when you multiply by any of the numbers 2, 3, . . ., $N-1$?

Well, 7 is certainly not unique in this respect. The first example other than 7 is provided by 17, which gives the decimal expression

1/17 = 0.0588235294117647 0588235294117647. . . . . .

a pattern which is simply shifted along when you multiply by any one of the numbers 2, 3, . . ., 16. Other examples less than 100 are 19, 23, 29, 47, 59, 61 and 97. More are known, though it has not been demonstrated that there is an infinite number of such examples.

So just what is it about these numbers which gives rise to this curious behaviour? Before there is any hope of answering that, we should ask ourselves just which numbers $N$ have reciprocals whose decimal expression consists of an infinite repeating pattern? This question is easily answered. A number $N$ will have a reciprocal whose decimal expression is finite if, and only if, the only prime factors of $N$ are 2 and 5. In all other cases (that is, if $N$ is divisible by a prime other than 2 and 5), $1/N$ will have an infinite repeating decimal expression. The following simple little argument demonstrates this fact.

---

*This chapter does get a bit technical, but if you can stomach it there is the reward of a nice problem to try out on your micro.

First of all suppose $N$ is of the form $2^r.5^s$ for some numbers $r$ and $s$ (possibly zero). Then

$$\frac{1}{N} = \left(\frac{1}{10}\right)^r \times \left(\frac{1}{5}\right)^{s-r}$$

or

$$\frac{1}{N} = \left(\frac{1}{10}\right)^s \times \left(\frac{1}{2}\right)^{r-s}$$

depending upon whether $r < s$ or $s < r$. (If by chance $r = s$, then $N$ is just a power of 10, in which case we know that $1/N$ is a finite decimal, so there is no need to consider this case at all.) But 1/2 and 1/5 are finite decimals, and when you raise them to any power the result will still be a finite decimal. And multiplying a finite decimal by a power of 1/10 also gives you a finite decimal. So $1/N$ will have a finite decimal expression.

Having shown that if $N$ has only 2 and 5 as prime factors, then $1/N$ is a finite decimal, we now try to prove the converse, namely that if $1/N$ is a finite decimal, then $N$ has no prime factors other than 2 and 5. Let

$$1/N = 0.a_1a_2...a_k$$

that is, let $a_1, ..., a_k$ be the digits of $1/N$, expressed in decimal form. Multiplying by $10^k$ you get

$$10^k/N = a_1a_2...a_k$$

Rearranging this gives

$$10^k = N \times a_1a_2...a_k$$

This implies that any prime factor of $N$ has to divide into $10^k$. But the only prime factors of $10^k$ are 2 and 5, so this does what we set out to do.

If a number $1/N$ does have an infinite decimal expression (that is, if $N$ does have a prime factor other than 2 and 5), then this decimal must eventually start to repeat itself. It may do so from the outset, such as

$1/3 = 0.3\,3\,3\ldots$

or else after one or more places, such as

$1/6 = 0.1\,6\,6\,6\ldots$

Or the repeating pattern may consist of more than one digit, as in the case of $1/7$.

If $1/N$ does consist of a repeating pattern, the maximum number of digits in the pattern is $N-1$. To see this (and at the same time show that the decimal does indeed cycle), consider the standard method for calculating $1/N$ by 'long division' of $N$ into 1. After each stage in this process, you are left with a remainder less than $N$. If this remainder is ever 0, the decimal expression for $1/N$ will terminate. As soon as the remainder term is repeated, from then on the entire division process will be repeated. But, except for 0, there are exactly $N-1$ possible remainders, so after at most $N-1$ steps the division process has to repeat itself.

If you want to follow this chapter any further, you need to know about the mathematical notion of congruence. So if you have not yet read chapter 8, you had better do so now.

Ready? If $N$ is such that $1/N$ is an infinite repeating cycle which is simply shifted along when you multiply by each of $2, \ldots, N-1$, then in order to get the $N$ distinct starting points for each of the expressions involved the length of the cycle involved has to be $N-1$, the maximum possible. This certainly implies that $N$ has to be a prime number, because if $k$ was a number between 1 and $N$ which divides into $N$, then $k/N$ would have a cycle length of at most $N/k - 1$, so it could not be a shift of $1/N$. But there are plenty of primes which do not have this property (such as 3, 5, 11, 13, 31). In fact, there is a good exercise to try out on your micro. Write a program to calculate, for each number $N$ from 1 to 100 (say), the cycle length of $1/N$. (Call the cycle length of a finite decimal expression 0, so that you get a numerical output for every number.)

It turns out that the above property, that $1/N$ has an infinite repeating decimal with a cycle of length $N-1$, is exactly the condition which characterises the numbers $N$ having the 'seven property' that we are looking at. To prove this fact, what we have to do is demonstrate that if $1/N$ does have such an expression, then $N$ has the seven property; that is, for each $k$ less than $N$, $k/N$ is just a shift of $1/N$. This is where congruences come into the picture.

Now, if $k/N$ is a shift of $1/N$ by $r$ places to the left, then the decimal parts of $k/N$ and $10^r/N$ must be the same, so $10^r/N - k/N$ will be a whole number. In other words, $N$ must divide exactly into $10^r - k$. Conversely, if for some number $r$, $N$ does divide exactly into $10^r - k$, then the expression for $k/N$ must be a shift of $1/N$ by $r$ places to the left. So what we have just discovered is that $k/N$ will be a shift of $1/N$ if, and only if, there is a number $r$ such that $N$ divides exactly into $10^r - k$. Or, in congruence language, $k/N$ will be a shift of $1/N$ if, and only if, there is a number $r$ such that

$$10^r \equiv k \pmod{N}$$

This last observation means that what we have to convince ourselves of is that if $1/N$ has an infinite repeating expression with a cycle of length $N-1$, then for any $k$ less than $N$ there is a number $r$ such that $10^r \equiv k \pmod{N}$. Actually, we do not need to bother about $k = 1$ here, since we know that $1/N$ is a 'shift' of $1/N$. But as it happens the result is true for $k = 1$. Why? Well, since $1/N$ has a cycle of length $N-1$, $N$ has to be prime. (If $N = P \times Q$, you can work out $1/N$ by first working out $1/P$ and then dividing the result by $Q$, using long division on both occasions. So the maximum cycle length for $1/N$ is $(P-1)(Q-1)$, which is less than $N-1$.) So by Fermat's Little Theorem (quoted in chapter 8, which I warned you to read!), $10^{N-1} \equiv 1 \pmod{N}$.

Having just observed that $10^{N-1} \equiv 1 \pmod{N}$, the next step is to show that if $r$ is less than $N-1$, then $10^r$ cannot be congruent to 1 modulo $N$. Well, suppose that there were some $r$ less than $N-1$ with $10^r \equiv 1 \pmod{N}$. Thus for some number $q$ it must be the case that

$$10^r - 1 = N \times q$$

Since $q$ can have at most $r$ digits, we can write $q = a_1 a_2 . . . a_r$, where each of $a_1, . . ., a_r$ is a single digit. Thus

$$\frac{1}{N} = a_1 a_2 . . . a_r \times \frac{1}{10^r - 1}$$

Next we make use of the formula

$$x + x^2 + x^3 + . . . = x/(1-x)$$

for the sum of an infinite geometric series, where $0 < x < 1$. If you put $x = 1/10^r$ in this formula you get

$$1/(10^r - 1) = (1/10^r)/(1 - 1/10^r) = 1/10^r + 1/10^{2r} + 1/10^{3r}\ldots$$

So

$$\frac{1}{N} = 0.a_1a_2\ldots a_r \quad a_1a_2\ldots a_r \quad a_1a_2\ldots a_r \ldots\ldots\ldots$$

This contradicts the fact that $1/N$ has a cycle of length $N-1$. So there cannot be such a number $r$ as postulated above.

Now that we know that $10^r$ is not congruent to 1 modulo $N$ for any $r$ less than $N-1$, it is an easy matter to show that if $r < s \leqslant N-1$, then $10^r$ cannot be congruent to $10^s$ modulo $N$. For suppose, on the contrary, that it were the case that

$$10^r \equiv 10^s \pmod{N}$$

Rearranging this congruence we get

$$10^r.(10^{s-r} - 1) \equiv 0 \pmod{N}$$

Since $N$ is prime, this means that either $N$ divides $10^r$ or else $N$ divides $10^{s-r} - 1$. But $N$ is neither 2 nor 5, so $N$ cannot divide $10^r$. And since $s-r < N-1$, $N$ cannot divide $10^{s-r} - 1$, by what was just proved above. So we have a contradiction, which means that $10^r$ and $10^s$ have to be non-congruent modulo $N$, as initially claimed.

OK! If you divide $N$ into anything and there is a remainder, that remainder has to be one of the numbers $1, 2, 3, \ldots, N-1$. In other words, any number not divisible by $N$ has to be congruent modulo $N$ to one of the numbers $1, 2, \ldots, N-1$. But by what was proved in the previous paragraph, no two of the numbers $10^1, 10^2, \ldots, 10^{N-1}$ leave the same remainder on division by $N$. So the numbers $10^1, 10^2, \ldots, 10^{N-1}$ have to be congruent modulo $N$ to each of the numbers $1, 2, \ldots, N-1$ *in some order*. (It does not matter what this order is.) If you think about it a moment, you will realise that this answers our question; that is, it shows that any number between 1 and $N-1$ (inclusive) is congruent modulo $N$ to a number of the form $10^r$.

So now you know. A number $N$ will have the 'seven property' if, and only if, the decimal expression for $1/N$ is an infinite repeating pattern of cycle length $N-1$. If you want to use your micro to find examples of such numbers, the best way is to make use of one of the results proved above. Namely, any such $N$ will be prime and have the property that $10^r$ is not congruent to 1 modulo $N$ for any $r$ less than $N-1$. It is easy to check this last condition for a given prime $N$ using a micro. In fact it can be made even easier. All you need to do is to check this condition for numbers $r$ which are factors of $N-1$. (It is not hard to prove that this is enough, but I think you have seen plenty to be going on with, and the more inspired among you might like to try this bit yourselves.)

### *Brain Power*

By general acknowledgement, Wim Klein (now aged 72) from Amsterdam is the fastest (human) computer. Klein's numerical skills caused him to be employed for many years as a 'computer' at the CERN atomic research laboratories in Geneva. Towards the end of 1983, Klein established two new world records which will be hard to beat.

For the first of his records, both of which were established in front of large audiences in a lecture hall at the DESY atomic research laboratories in Hamburg, Klein factored a 4 digit number into a sum of four squares in 10 different ways. It took him, in all, 43.8 seconds. The date was 30 September 1983. The number in question was 5462, and was chosen by the audience. One of the 10 factorisations is

$$5462 = 73^2 + 9^2 + 6^2 + 4^2$$

How long will it take you to find the other 9, even using a calculator? Come to that, how long will it take a computer, given a suitable program in store?

On 22 November 1983, the audience picked a 505 digit number from a list of 42 possibilities calculated on the DESY computer, all of which were 73rd powers. In 1 minute and 43 seconds, Klein was able to calculate the 73rd root of this number.

How are such feats possible? Well, for most of us they are not, of course, but there are some tricks that Klein employs. For instance, he knows all the four-figure logarithms in base 10 up to 150 off by heart, and by being able to split any number greater than 150 into prime factors extremely quickly he can extend his use of logarithms to numbers beyond 150. And for calculating the 73rd root he used methods which required knowledge of only the first five digits and the last three (29326 and 344, respectively, if that is any help to you), so at least there was no need to memorise the entire number! Did I just call these 'tricks'?

# *13 Primes and secret codes*

Late in 1982, experts in cryptography, the science of secret codes, were stunned to discover that a coding method which was widely regarded as the most secure in the world was not, as it stood, as secure as was thought. A remark made over a beer in a bar in Winnipeg, Canada, concerning the rather eccentric mathematical problem of how you factor numbers using a computer, sent the cryptographers scurrying for their reference manuals to amend their codes before the 'enemy' heard the news. Though the basic idea behind the method (known as the RSA System, after the mathematicians Ronald Rivest, Adi Shamir and Leonard Adleman, who invented the system) remained sound (it is believed), the actual implementation of the method had to be changed.

The RSA System is an example, in fact the most 'popular' example, of a coding method known as *Public Key Cryptography*. To understand what this system involves, you need to know a little bit about how any modern coding system works. Suppose I wanted to send you a message, and that it was important to us that no third party could intercept and obtain the message. Assuming that we were some distance apart, possibly even continents apart, we would have to use some kind of electronic method of communication. (I suppose we could use some form of letter, but that would be hopelessly slow, and in general impractical, which also cuts out the possibility of the obviously secure method of using a trusted courier. So let us stick to the idea of using fast, electronic means.) With the eavesdropping hardware available nowadays, there is no way we can stop an enemy actually picking up our communication, so our only hope for security is to scramble the message prior to sending, and unscramble it at the receiving end. For this to work, the scrambling should be sufficiently complicated to resist attempts by an enemy to figure it out and yet simple and regular enough for the scrambling/descrambling to be performed with reasonable efficiency. In particular, you, the receiver, need to know how to unscramble the seemingly meaningless message that you actually receive.

The problem with any suitable coding system is obvious. How do you communicate to the potential message-receiver the secret information required to decode the message? One method is to use a trusted courier *to set up the coding system*, and then use it for some length of time. This has two disadvantages. Firstly, if a large network of communicators is involved, the poor couriers end up travelling all over with enormous amounts of information. (A secondary problem is that the courier may be bribed to disclose the secret, but let us be charitable and ignore this point for the moment.) The second disadvantage is that the longer any coding system is in use, the more liable it is to being broken by an enemy, who can be assumed to be very clever and have available immense computing power. (The details of just how codes can be broken is a fascinating subject in its own right, by the way, but that is another story.)

A common method of coding is to design a computer program which permutes the components of a message in some complicated way depending upon a numerical parameter or 'key'. Thus, when the system has been used for some time, and is thus less secure, all that is required is to change the numerical 'key'; there is no need to design another algorithm afresh. An example of such a system which is in widespread use is the DES (*Data Encryption Standard*). This uses a key which is a 56 bit number (in binary form). In order to try to find the key by a trial and error procedure, the enemy is faced with looking at $2^{56}$ possible keys, a truly astronomical number (see chapter 1). It is the large number of possible keys which gives the DES its security, though it has been argued that with present day computer speeds, a number of this magnitude is not really far enough out of range. At any rate, with the DES and similar systems, decoding the message is the reverse of encoding, and uses the same key (only 'backwards'). So there is still the problem of sending the key to a potential receiver.

In 1975, Whitfield Diffie and Martin Hellman proposed a revolutionary new type of encryption method: *Public Key Cryptography*. The idea here, as with more conventional systems, is to use a computer program to encrypt and decrypt messages, encryption and decryption both depending upon a numerical key. But instead of the encryption key and the decryption key being the same, as is the case with conventional systems, with a Public Key System they are quite different. The encryption key is made public (hence the name of these systems), only the decryption key being kept secret. It is as if you needed two keys to your house, one to lock the door and another to unlock it. Clearly it would be a very unusual lock which required different keys for locking and unlocking, and likewise it is an unusual encryption scheme which requires different keys for encryption and decryption.

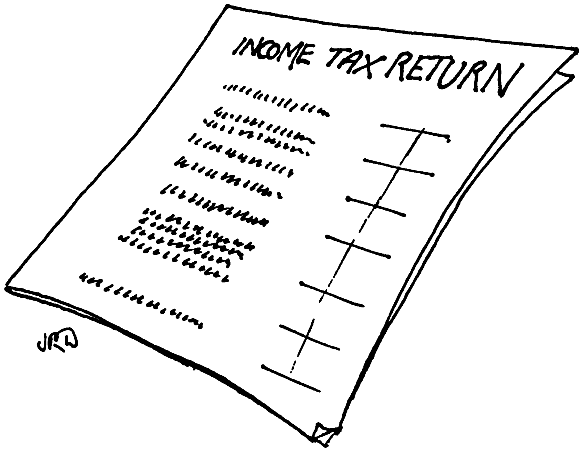
Users of a Public Key System proceed as follows. Each user chooses his own secret decryption key. This number is never revealed to anyone. To avoid detection by trial and error searches, it will be a very large number, possibly with more than 100 digits. From this number, he calculates, in an entirely standard fashion, his own encryption key, which is listed in a public directory of users of the system. When system user A wants to send an encoded message to user B, he first looks up B's public key. Using this key, A encrypts the message. When B receives the encrypted message from A, he uses his secret decryption key to decode it. Since only B knows the decryption key, only he can carry out this decoding procedure. Not even A, who sent the message, can decode it! (So A had better not forget what it was he originally said in the message.)

There are several ways of setting up a Public Key System of the above kind. The most common one at the moment is the RSA system, mentioned at the outset. Without going into too much detail, a user of this system will choose two large prime numbers for the secret decoding key. The public encoding key is the product of these numbers. Encryption of a message corresponds to the multiplication of these two primes (a simple task), decryption to factoring the product number into its component primes (a task for which no practicable method is known). (In practice things are not quite as simple as this, but this should convey the essence of the thing.) The security of this system depends upon the fact that there is no efficient method for factoring large numbers. A potential enemy is more or less reduced to trying to find the factors by trial and error. So, providing that the two primes used initially as the decryption key are large enough, such a system should be secure.

Using modern primality tests (see chapters 1 and 8), it takes a fast computer at most 15 seconds to check if a 50 digit number is prime, and no more than 40 seconds for a 100 digit number. Factorisation, on the other hand, is a much more difficult problem. Until very recently, the best results involved numbers of at most 50 digits or thereabouts. Even a very fast computer would require many years to factor a number much larger than this. Consequently, early users of the RSA encryption system made use of two 50 digit primes in their codes. (For efficiency of use, it is better to use primes which are as small as security allows. The bigger the keys, the more computation is required in encryption and decryption.)

Though there are some clever methods for factoring numbers (indeed, without using mathematically based approaches, even 30 digit numbers may prove impossible to factorise), mathematicians believe that there is no efficient method which would compare with the efficient primality tests. Consequently, attention has turned to the hardware involved in factorisation. (As you can imagine, with a subject involving security codes, there is no shortage of money for research!) For instance, in 1982, Marvin Wunderlich, a factoring expert from America, spent some months in Britain working at Queen Mary College, London, trying to develop a factoring program on a computer known as the DAP (*Distributed Array Processor*). This unusual machine, built by ICL, has some 4096 arithmetic processing units working simultaneously, which makes it very fast when suitably programmed. Unfortunately, Wunderlich had to leave for home before he could get the program running properly. But the lesson had been learnt. On his return to the USA, Wunderlich resumed work on a factoring program working on a CRAY-1 computer. Though not a 'parallel computer' in the sense of the DAP, the CRAY-1 can handle long strings of computer words (that is, numbers) more or less at the same time, which is what is required in factoring algorithms.

Then came the legendary beer. At a scientific meeting in Winnipeg, Wunderlich and another factoring expert called Gus Simmons went out for a beer with Cray engineer Tony Warnock. The discussion turned to the problem that the two factorers were facing in trying to speed up their method: namely, the necessity to keep making very small changes to long lists of numbers. Warnock at once pointed out that the way the CRAY-1 handled long strings of numbers meant that this could be achieved in a very short space of time, by-passing a lot of the work Wunderlich and Simmons' existing program was doing. The result of this chance remark? Wunderlich and Simmons were soon able to factorise 60 digit numbers in about one hour, and were already setting their sights on 70 digit numbers using computers built specially for the job. RSA users are now advised to employ 100 digit primes in their codes. Factoring 200 digit numbers still seems safely out of range. At least for the time being!


INCOME TAX RETURN

THE WORLD'S MOST UNBREAKABLE SECRET CODE

## *Fermat's Factorisation Method*

Suppose you try to factorise a fairly large number into a product of two smaller numbers. The obvious approach is to keep trying all the prime numbers, one after another, seeing if any of them is a factor of the number. Such a method will succeed only if the number involved *does have* a 'small' factor. But what if the number is a product of two 10 digit primes? Then an approach such as this would be very inefficient. For much larger numbers it would be totally infeasible. Modern factoring algorithms begin by looking for small factors, but if none are found they switch to more sophisticated methods to find large factors. The basic idea behind these other methods goes back to the great seventeenth century mathematician Pierre De Fermat.

Fermat's factorisation method works like this. Suppose that you want to factorise a number $n$ into two factors $a$ and $b$, which are roughly equal in size. You make the observation that this problem is equivalent to finding two numbers $x$ and $y$ such that

$$n = x^2 - y^2$$

For if you have $x$ and $y$ like this, you get $a = x + y$ and $b = x - y$, making use of the algebraic identity

$$x^2 - y^2 = (x + y)(x - y)$$

Conversely, if $n = a \times b$, then (provided that $n$ is odd, which we may assume from the outset since even numbers are recognised at once and factors of 2 can be divided out) $x = (a + b)/2$ and $y = (a - b)/2$ will give

$$n = x^2 - y^2$$

The idea now is to find numbers $x$ and $y$ with the above property, or what is the same thing, such that

$$x^2 - n = y^2$$

What you do is to start with the smallest value of $x$ for which $x^2 - n$ is positive, and then keep increasing $x$ by 1 until $x^2 - n$ works out to be a perfect square, in which case $y$ is its square root and the number $n$ is factorised. This can be speeded up by using special tricks to see if a number is a perfect square or not. For instance, only numbers which end in 0, 1, 4, 5, 6, 9 can be perfect squares, so numbers ending in 2, 3, 7, 8 can be disregarded at once.

Fermat himself used the above method to factorise the number

2027651281

into a product of two roughly equal primes. Why don't you try it yourself? (Remember, Fermat had no calculating aids, so you should not use any either.)

Then try the numbers 119143 and 340663. And when you have got this far, how about trying to write a factorisation program for your micro? Your program should begin by looking for small factors, and if none are found should switch to Fermat's method to find large factors. Look for tricks to speed things up. If you find your program can factor 70 digit numbers, try to sell it to the highest bidder and become rich (or imprisoned) overnight.

# *14 Perfect numbers*

A *perfect* number is one which is equal to the sum of its proper divisors. For example, the divisors of the number 6 are (excluding 6 itself), 1, 2, 3, and

$$1 + 2 + 3 = 6$$

so 6 is perfect. The next example of a perfect number is 28, with

$$28 = 1 + 2 + 4 + 7 + 14$$

These two examples were known to the Pythagoreans, who gave them the name of 'perfect' numbers. It was thought that such numbers had mystical significance. For instance, the Old Testament gives 6 days as the time required for the Creation, and the Moon takes 28 days to circle the Earth.

After 28, the next two perfect numbers are

496 and 8128

These were known by the end of the first century A.D., and are cited in the manuscript *Introductio Arithmeticae* by the Greek mathematician Nicomachus. Demonstrating how misleading numerical evidence can be in mathematics, Nicomachus' conjectures that the $n$th perfect number would contain $n$ digits and that the perfect numbers would end alternately in 6 and 8 were both refuted by the discovery, in the fifteenth century, of the fifth perfect number

33,550,336

and in the sixteenth century, of the sixth

8,589,869,056

It is, however, the case that any even perfect number will end in either 6 or 8; indeed, either 6 or 28, the first two perfect numbers.

The appearance in the previous paragraph of the adjective 'even' for perfect numbers may be unnecessary. It is widely believed that all perfect numbers are even, though this has not been proved. Computer aided work has demonstrated that if an odd perfect number was to exist it would have to exceed $10^{100}$ and have at least 11 distinct prime factors. In view of which, let us agree that for the rest of this chapter 'perfect number' will mean '*even* perfect number'.

In Book IX of the famous *Elements* (about 350 B.C.), Euclid showed that if the number $2^n - 1$ is a prime for any $n$, then the product $2^{n-1}(2^n - 1)$ is perfect. For instance, for $n = 2$, $2^n - 1 = 3$, which is prime, so $2^{n-1}(2^n - 1) = 2 \times 3 = 6$ is perfect. Some 2000 years were to elapse before Euler showed that the converse to Euclid's result is true: namely, if $P$ is a perfect number, then $P$ will be of the form $2^{n-1}(2^n - 1)$ for some number $n$ such that $2^n - 1$ is prime. Consequently, looking for perfect numbers is equivalent to looking for prime numbers of the form $2^n - 1$. And there, totally unexpected, is one of those rare but wonderful occasions in mathematics when two apparently unrelated notions turn out to coincide. Primes of the form $2^n - 1$ occurred way back in chapter 1 of this book; they are the *Mersenne primes*. Since there are, to date, precisely 29 Mersenne primes known, we know at once that there are just 29 known perfect numbers. We also know that they are, for the most part, enormously large.

Speaking of the largeness of perfect numbers, in his 1811 book *Theory of Numbers*, Barlow, after giving the eighth perfect number, the 19 digit number $2^{30}(2^{31} - 1)$, wrote that it "is the greatest that ever will be discovered; for as they are merely curious, without being useful, it is not likely that any person will ever attempt to find one beyond it." As we now know, human curiosity being what it is, we have come a long way since Barlow made this rash statement.

The perfect numbers have some curious numerological properties. For instance, every perfect number is *triangular*, that is, it can be realised as the number of balls arranged to form an equilateral triangle (which is equivalent to being of the form $\frac{1}{2}n(n + 1)$ for some number $n$). Again, the sum of the digits of any perfect number other than 6 always leaves a remainder of 1 on division by 9. Related to this is the fact that the *digital root* of any perfect number other than 6 is 1. The *digital root* of any number is obtained by adding together all the digits in the number, then adding together all the digits in that sum, then the digits in that number, and so on, until a single digit number is reached. For instance, for the fifth perfect number, 33,550,336, adding the digits gives

$$3+3+5+5+0+3+3+6=28$$

and adding the digits of 28 gives 10, and adding the digits of 10 gives 1.

Still more odd facts. Every perfect number is the sum of consecutive odd cubes. For instance

$$28 = 1^3 + 3^3$$

$$496 = 1^3 + 3^3 + 5^3 + 7^3$$

And if $n$ is perfect, then the sum of the reciprocals of all the divisors of $n$ is always equal to 2. For instance, 6 has divisors 1, 2, 3, 6, and

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 2$$

None of the above facts is particularly difficult to prove, though unless you are fairly adept at algebraic manipulations they could cause you a few headaches. You can find further details in my book *Microchip Mathematics: Number Theory for the Computer User* (Shiva Publishing, 1984).

The sheer size of perfect numbers puts the problem of finding new ones out of range of the home micro user, but there are some related problems that are still solvable. For instance, two numbers are said to be *amicable* if each is equal to the sum of the proper divisors of the other. For example, the numbers 220 and 284 form an amicable pair, a fact discovered by Pythagoras around 500 B.C. The relevant calculations are

$$1+2+4+5+10+11+20+22+44+55+110=284$$

(adding together the proper divisors of 220)

$$1+2+4+71+142=220$$

(adding together the proper divisors of 284).

To date, about 3000 amicable pairs are known. Why don't you see how many you can find. If you find more than 3000, let me know!

A more general notion is this. By a *sociable chain* of numbers is meant a finite sequence of numbers such that each one is the sum of the proper divisors of the preceding one, with the last number in the chain regarded as preceding the first. Very few examples of sociable chains are known. There is one consisting of five numbers each between 12,000 and 16,000. When you have found this, you could try to find some new ones. Let me know of any success you have, and you will be rewarded with the rare honour of appearing in future editions of this tome.

LOOKING FOR PERFECT NUMBERS

### *Bugs*

'Debugging' is one of the first words people learn when they start to write computer programs. It is well nigh impossible to write a program of more than half a dozen lines (if that) without some small and practically undetectable error creeping in which, inevitably, causes the program to produce mysteriously bizarre results when it is run (always assuming that it will run at all). 'Bugs' are what they are called, and 'debugging' is the art (if that is the word) of finding and removing them.

It has been estimated that around 80 per cent of all the time (and hence money) spent on writing commercial software is spent on debugging. And it is unlikely that any large-scale program is ever rendered totally bug-free. This goes for the operating systems and language compilers upon which everyone's programs ultimately depend, besides the programs written by computer users.

But why 'bugs' and 'debugging'? Believe it or not, the use of this terminology stems from the good old string and sealing-wax days of computing. In the 1940s at Harvard University they built a computer known as the Mark I Automatic Relay Calculator. A malfunction in one of the signal relays one day was found to be caused by the presence of a dead moth. And so the operator who removed it performed the first ever recognised 'debugging' operation. Rumour has it that the dead moth (or what is left of it) can still be seen, stuck to a page in a log book in the Navy museum in Virginia.

# *15 True beyond reasonable doubt*

Number Theory has some of the flavour of an experimental science. Numerical calculations may be used to gather evidence in favour of (or against) conjectures, and may in fact lead to the formulation of new conjectures. Computers provide the mathematician with an efficient tool for gathering this evidence. But you have to be careful. Mathematics does not take too kindly to being treated in an offhand manner, and has quite a few surprises in store for the unsuspecting computer user. Or for the unsuspecting mathematical genius for that matter. Even the great Fermat fell into the trap of extrapolating from numerical evidence to what turned out to be a false conclusion.

In honour of his mistake, numbers of the form

$$F_n = 2^{2^n} + 1$$

are nowadays called *Fermat numbers*. Fermat's mistake was to assume that all such numbers are prime. The evidence is particularly shaky. He observed that the first five such numbers are prime, namely

$$F_0 = 3, F_1 = 5, F_2 = 17, F_3 = 257, F_4 = 65{,}537$$

And on the basis of this observation he wrote to Mersenne that "I have found that numbers of the form $2^{2^n} + 1$ are always prime numbers and have long since signified to analysts the truth of this theorem." Unfortunately, he was unable to find a proof of his claim, though the evidence of his letters indicates that it was not for want of trying. In 1732, Euler showed why Fermat had been unsuccessful. The number

$$F_5 = 2^{2^5} + 1 = 4{,}294{,}967{,}297$$

is divisible by 641, and hence is not prime.

Subsequent work has shown that, contrary to Fermat's belief, the numbers $F_n$ are non-prime for all values of $n$ from 5 to 16, as well as for some other, isolated values of $n$. The sheer size of the Fermat numbers (except for the first few) makes such work extremely difficult, even using powerful computers. By and large, the results showing non-primality have made use of efficient primality tests (see chapter 8), which prove non-primality without giving any factors of the number. In the case of $F_7$, which written out in full is

$$340282366920938463463374607431768211457$$

this number was shown to be non-prime back in 1905, but it was not until 1971 that Brillhart and Morrison discovered the factorisation

$$F_7 = 59649589127497217 \times 5704689200685129054721.$$

(They used a computer, of course, but even then it required an entire new theory of factorisation techniques.)

The discovery (in 1953) that $F_{16}$ is not prime put to rest another widely held belief, namely that all the numbers in the infinite sequence

$$2 + 1, 2^2 + 1, 2^{2^2} + 1, \text{etc.}$$

are prime. In fact, the modern conjecture is that $F_n$ is not prime for any value of $n$ beyond 4, the opposite of Fermat's claim. But are modern mathematicians being as rash as Fermat in making such a claim?

Of course, though the numbers involved in the above are large, Fermat based his false claim on rather slender evidence: just five cases. What about more overwhelming evidence? For instance, the conjecture that all perfect numbers are even, mentioned in chapter 14, has been verified up to $10^{100}$, the Goldbach Conjecture that every even number is the sum of two primes has been checked up to 100 million, and Fermat's Last Theorem (see chapter 11) has been checked for exponents up to 125,000. Surely we can rely on evidence such as this? In fact we had better not. Two spectacular examples should serve to illustrate the danger of such a practice.

The first example concerns the way in which the prime numbers are distributed among the rest of the whole numbers. A classical theorem of Number Theory, called the *Prime Number Theorem*, says that for very large $n$, the number of primes less than $n$ is approximately equal to $n/\log(n)$, where $\log(n)$ is the natural logarithm of $n$. (This in itself is an amazing result, since it shows that there is a connection between the prime numbers and the number e, the base of the natural logarithms – see chapter 10 for more coincidences involving e.) The approximation involved here improves as $n$ gets bigger, and by choosing $n$ large enough the approximation can be made as accurate as you please. But there are other approximating functions which are even better. One of these is denoted by the symbols Li($n$). For any $n$, Li($n$) is a pretty good approximation to the number of primes less than $n$. For instance, Li(1000) = 178 and there are 168 primes less than 1000. Again, Li(1,000,000) = 78,628, and there are 78,498 primes less than 1,000,000. (The exact definition of Li($n$) is not important here, but for those who are able to understand the various notions involved, Li($n$) is the integral of the function 1/log($x$) from 2 to $n$.)

For any $n$, Li($n$) approximates the number of primes less than $n$ on the high side. Or so it appears. No-one has found a number $n$ for which the number of primes less than $n$ exceeds the value of Li($n$), despite some very long computer searches. (I am not sure how far such searches have been carried now, but it is certainly more than 10 billion.) This is overwhelming evidence that Li($n$) always exceeds the number of primes less than $n$, you might well think. Well, evidence or not, the mathematician J. E. Littlewood proved some decades ago that Li($n$) is sometimes less than the number of primes less than $n$. Not that Littlewood's proof provided an example of such a number $n$. The best that he could do was to show that such an $n$ would certainly occur before the number

10 raised to the power (10 raised to the power (10 raised to the power 34))

a number for which the adjective 'astronomical' would be totally inadequate. In fact, for many years, this number was the largest one that had ever played a significant role in mathematics. (I will not tell you what the new record holder is. At least, not yet.) This would seem to imply that we shall *never* know the value of a number $n$ for which Li($n$) is less than the number of primes below $n$, either with or without the aid of computers, even though we know there is such a number. (Or am I being too much of a defeatist?)

The second example of the numerical evidence being misleading involves a result proved as recently as the spring of 1984: using a combination of traditional mathematical reasoning and computer methods, Andrew Odlyzko of Bell Laboratories and Herman Te Riele of the Amsterdam Center for Mathematics and Computer Science showed that a long-standing conjecture due to T. J. Stieltjes (1885) and F. Mertens (1897), and now generally known as *Merten's Conjecture*, is false.

Merten's Conjecture is easily stated. For any number $n$, let $M(n)$ denote the difference between the number of numbers less than $n$ which are products of an even number of distinct primes and the number that are products of an odd number of distinct primes. (Numbers which are divisible by the square of a prime are thus ignored in calculating $M(n)$.) The conjecture is that $M(n)$ is always less than the square root of $n$. Merten's checked this for all values of $n$ up to 10,000 using pencil and paper techniques, a task which presumably led to the conjecture being named after him and not after Stieltjes who formulated it first. And in 1963, a computer search checked it up to 10 billion.

But, 10 billion cases or not, it is still false. As with the Littlewood result mentioned a few moments ago, the new result on Merten's Conjecture does not actually provide us with a value of $n$ for which $M(n)$ exceeds the square root of $n$. All we know is that there has to be such an $n$. Odlyzko and Te Riele themselves hazard the *guess* that there will be such an $n$ around

10 raised to the power (10 raised to the power 70)

which again would be out of reach of any computer that we could ever build.

Of course, you *could* try to prove me wrong and find an example on your home micro. The problem is easy to program, though with numbers greater than 10 billion you would have to write a very efficient program to stand even a remote chance of success. So, on reflection, perhaps it is wiser to content yourself with repeating Merten's calculation up to 10,000, and seeing just how close $M(n)$ gets to the square root of $n$. But, if you are determined to go ahead and venture into the unknown, then good luck to you, and, if you succeed, please remember who gave you the idea when you find yourself rich and famous as a result.

## ***Problems, Problems, Problems***

With only two exceptions, every odd number between 1 and 9000 is either a prime or else the sum of a prime and twice a square. What are the two exceptions?

******

What is the largest number whose prime factors add up to 100? (1 is not regarded as a prime factor here.)

******

In the following multiplication sum, each digit * is prime. Fill in all the digits.

```
    * * *
         X
      * *
---------
  * * * *
* * * * 0
---------
* * * * *
---------
```

******

What is it that is so unusual about this particular paragraph? I doubt if you could find any additional such paragraphs in this (or in practically any) book. To supply you with a hint: a cryptography buff would not find it hard at all to say what is so odd. It has to do with probability. But the probability of what? Your micro cannot bail you out on this occasion, it is all up to you.

******

There are only four numbers which are equal to the sum of the cubes of their digits. What are they?

# *16 All numbers great and small*

Some time around 760 A.D., an Arabian mathematician called *al-Khowarizmi* wrote a book outlining the rules for performing basic arithmetic using numbers expressed in the Hindu decimal form that we use today, with columns for units, tens, hundreds, etc., and decimal points to denote fractions. From his name comes the modern word 'algorithm', nowadays used to describe any set of rules for performing calculations. Computer programs are just algorithms expressed in some computing language such as BASIC, FORTRAN, or PASCAL. (At least, computer programs concerned with mathematical calculations are algorithms. In other cases it may or may not be the case that an 'algorithm' is involved.)

The simplest numerical algorithm is the method of addition that we all learn in school. To add two numbers, we write one under the other, aligned in columns starting from the right hand side, and add the individual digits, column by column, starting at the right, with carries whenever a multiple of 10 occurs. This set of instructions may be applied for any pair of numbers, however large. (Of course, I have stated the algorithm extremely briefly, since I assume any reader of this book is familiar with the method.) The same algorithm may be used to enable computers to handle numbers larger than the computer's word size. The idea is to split the numbers up into pieces which will fit into a computer word. (In order to allow for multiplication routines, it is in fact more usual to split up the numbers so that each piece fits into *one-half* of a computer word.)

For example, suppose that you are using a 16-bit computer. The largest number which will fit into one word is then of the order of 16,000. Now, if you take two numbers less than 100 and multiply them together, the result will not cause overflow. So it would be convenient to split your numbers up by pairs of digits, so that, for example, the number 137890036 would be written as

1 37 89 00 36

This would be stored in five computer words. To add two numbers stored in this way, you would add them word by word, starting from the right. After each individual addition, any third digit which appears is carried to the left (that is, it is stored to be added on at the next stage), only the two rightmost digits of the sum being stored in the answer array. (So after each individual addition, you need to calculate the quotient and remainder on division by 100.) This is just the same as the school method for addition mentioned earlier, except that instead of working in multiples of 10 the computer works in multiples of 100.

Multiplication is performed similarly. That is, you take the standard method learnt in school, and perform it on the computer with 100 as the base instead of 10. The array facilities provided by most computer languages make this a relatively easy task to perform, though you may have to decide in advance the maximum length of number that you will need to handle, and must thereafter remember to 'fill in' the left hand part of smaller numbers with zeros.

There are various ways of speeding up arithmetic algorithms of this kind. For a start, since computers perform their arithmetic in binary form, it is much more efficient to write the routines in this manner, splitting the numbers by powers of 2 rather than powers of 10 as described. But since this inevitably means that you have to write the routines in some form of assembly language, it is only sensible if you intend to perform very complicated calculations on many occasions. For occasional use, ordinary decimal routines should suffice. But the experts do not stop at simply writing binary routines. There are some extremely clever tricks employed to make multiplication, in particular, much faster than might at first be thought possible.

One method is this. Suppose that $u$ and $v$ are two numbers which are of the same length in binary notation, each having $2n$ bits. (This will cover all cases, since there is nothing to stop us using leading zeros to pad out shorter numbers.) Let the binary representations of these numbers be

$$u = u_{2n} \ldots u_n \ldots u_1 \quad \text{and} \quad v = v_{2n} \ldots v_n \ldots v_1$$

(So each of $u_1, \ldots, u_{2n}, v_1, \ldots, v_{2n}$ is either 0 or 1.) Split the numbers $u$ and $v$ into equal halves like this

$$U_L = u_{2n} \ldots u_{n+1},\ U_R = u_n \ldots u_1,\ V_L = v_{2n} \ldots v_{n+1},\ V_R = v_n \ldots v_1$$

By elementary arithmetic, the following equation is valid

$$uv = (2^{2n} + 2^n)U_L V_L + 2^n(U_L - U_R)(V_R - V_L) + (2^n + 1)U_R V_R$$

Take a close look at the right hand side now. Only three multiplications of pairs of $n$-bit numbers are involved. The rest consists of additions, subtractions, and 'multiplications' by powers of 2 (which in binary arithmetic are not really multiplications at all, of course, just 'shifts' to the left). The traditional method of multiplication using 'cross multiplication' would require four $n$-bit multiplications, namely

$$uv = 2^{2n} U_L V_L + 2^n (U_L V_R + U_R V_L) + U_R V_R$$

So the effect of using the former method is to replace one $n$-bit multiplication by some additions. Since addition of large numbers can be performed much faster than multiplication on a computer, this represents a net saving of computer time.

How do you multiply the various $n$-bit numbers that you end up with in this manner? Well, if these numbers are small enough to fit into half a computer word, you just use the multiplication hardware of the computer. If they are longer, you split them into halves and use the same trick again to reduce them. And so on. The result of this is that, while a naive multiplication algorithm will require a running time proportional to $n^2$ to multiply two $n$-bit numbers, the method outlined here only uses a time proportional to $n^{1 \cdot 6}$, which for large $n$ can be a real saving.

But this is only the beginning of the story. There are other algorithms. The fastest multiplication algorithm that I know of is due to Schönhage and Strassen. This algorithm is much too complicated to describe here, but let me just tell you that it allows you to multiply two $n$-bit numbers in a time proportional to $n \times \log(n) \times \log\log(n)$. Since the logarithm function grows incredibly slowly, this means that, over any reasonable range, the running time depends only upon $n$, a phenomenal saving against the naive method with running time depending upon $n^2$. If you want to discover more about this kind of thing, you should have a look at the book *The Art of Computer Programming, Part 2*, by Donald Knuth. However, I should warn you, this is a book written for the professional computer programmer with a sound mathematical background, so be prepared for a hard (but ultimately highly rewarding) slog. A slightly gentler account of computer arithmetic (but one which does not include the Schönhage–Strassen algorithm) can be found in my book *Microchip Mathematics: Number Theory for the Computer User.* (Knuth's book is published by Addison–Wesley, mine by Shiva Publishing.)

Table of the Mersenne primes known in June 1984

| | $n$ | Size of $2^n - 1$ | Date of discovery | Discovered by |
|---|---|---|---|---|
| 1. | 2 | 3 | antiquity | known to Euclid |
| 2. | 3 | 7 | antiquity | |
| 3. | 5 | 31 | antiquity | |
| 4. | 7 | 127 | antiquity | |
| 5. | 13 | 8191 | 1461 | unknown |
| 6. | 17 | 131071 | 1588 | Pietro Antonio Cataldi |
| 7. | 19 | 524287 | 1588 | |
| 8. | 31 | 2147483647 | 1772 | Leonhard Euler |
| 9. | 61 | 19 digits | 1883 | I. M. Pervouchine |
| 10. | 89 | 27 digits | 1911 | R. E. Powers |
| 11. | 107 | 33 digits | 1914 | R. E. Powers and E. Fauquemberge |
| 12. | 127 | 39 digits | (1876) 1914 | (Edouard Lucas) E. Fauquemberge |
| 13. | 521 | 157 digits | 1952 | Raphael Robinson (+ SWAC computer) |
| 14. | 607 | 183 digits | 1952 | |
| 15. | 1279 | 386 digits | 1952 | |
| 16. | 2203 | 664 digits | 1952 | |
| 17. | 2281 | 687 digits | 1952 | |
| 18. | 3217 | 969 digits | 1957 | Hans Riesel (+ BESK computer) |
| 19. | 4253 | 1281 digits | 1961 | Alexander Hurwitz (+ IBM-7090) |
| 20. | 4423 | 1332 digits | 1961 | |
| 21. | 9689 | 2917 digits | 1963 | Donald Gillies (+ ILLIAC-II) |
| 22. | 9941 | 2993 digits | 1963 | |
| 23. | 11213 | 3376 digits | 1963 | |
| 24. | 19937 | 6002 digits | 1971 | Bryant Tuckermann (+ IBM 360-91) |
| 25. | 21701 | 6533 digits | 1978 | Laura Nickel and Curt Noll (+ CDC-CYBER-174) |
| 26. | 23209 | 6987 digits | 1979 | Curt Noll (+ CDC-CYBER-174) |
| 27. | 44497 | 13395 digits | 1979 | Harry Nelson and David Slowinski (+ CRAY-1) |
| 28. | 86243 | 25962 digits | 1982 | David Slowinski (+ CRAY-1) |
| 29. | 132049 | 39751 digits | 1983 | David Slowinski (+ CRAY-XMP) |

# CRIB

The answers to the questions posed at the end of chapter 1 are

1. 1/3 = 5823/17469.
2. 1/4 = 7956/31824; 1/5 = 2973/14865; 1/6 = 2943/17658;
   1/7 = 5274/36918; 1/8 = 9321/74568; 1/9 = 8361/75249.
3. 148/296 + 35/70 = 1.
4. 15.
5. 25 and 52.

The answer to the question posed at the end of chapter 4 is

121 , 10000

The sequence given lists the expressions for the number sixteen in bases 16, 15, 14 down to base 4, and what you have to do is to supply the expressions for sixteen in bases 3 and 2. Simple, isn't it?